



Lenovo NeXtScale System Planning and Implementation Guide

**Introduces the high density x86
solution for scale-out
environments**

**Covers the air-cooled NeXtScale
System offerings: nx360 M5 & M4
nodes, and n1200 enclosure**

**Addresses power, cooling,
racking, and management**

**Provides the information you need for
a successful implementation**

David Watts

Matt Archibald

Jerrold Buterbaugh

Duncan Furniss

David Latino





Lenovo NeXtScale System Planning and Implementation Guide

August 2015

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

Last update on August 2015

This edition applies to:

NeXtScale n1200 Enclosure, type 5456
NeXtScale nx360 M5 Compute Node, type 5465
NeXtScale nx360 M4 Compute Node, type 5455
42U 1100mm Enterprise V2 Dynamic Rack, 93634PX
Rear Door Heat eXchanger for 42U 1100 mm Enterprise V2 Dynamic Racks, 1756-42X

© **Copyright Lenovo 2015. All rights reserved.**

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

Contents

Notices	vii
Trademarks	viii
Preface	ix
The team who wrote this book	x
Comments welcome	xii
Summary of changes	xiii
August 2015	xiii
July 2014	xiii
April 2014, Second Edition	xiv
Chapter 1. Introduction	1
1.1 Evolution of the data center	2
1.1.1 Density	2
1.1.2 Scale out applications	2
1.2 Summary of the key components	3
1.2.1 Lenovo NeXtScale n1200 Enclosure	4
1.2.2 NeXtScale compute nodes	5
1.3 Design points of the system	7
1.4 NeXtScale System cooling choices	8
1.5 This book	8
Chapter 2. Positioning	9
2.1 Market positioning	10
2.1.1 Three key messages with NeXtScale	11
2.1.2 Optimized for workloads	13
2.2 System x and ThinkServer overview	14
2.3 NeXtScale System versus iDataPlex	15
2.4 NeXtScale System versus Flex System	16
2.5 NeXtScale System versus rack-mounted servers	18
Chapter 3. NeXtScale n1200 Enclosure	19
3.1 Overview	20
3.1.1 Front components	22
3.1.2 Rear components	23
3.1.3 Fault tolerance features	24
3.2 Standard chassis models	24
3.3 Supported compute nodes	25
3.3.1 nx360 M4 node support	26
3.3.2 nx360 M5 node support	31
3.4 Power supplies	45
3.5 Fan modules	48
3.6 Midplane	50
3.7 Fan and Power Controller	51
3.7.1 Ports and connectors	51
3.7.2 Internal USB memory key	53
3.7.3 Overview of functions	53
3.7.4 Web GUI interface	54

3.8	Power management	55
3.8.1	Power Restore policy	55
3.8.2	Power capping	56
3.8.3	Power supply redundancy modes	56
3.8.4	Power supply oversubscription	56
3.8.5	Acoustic mode	58
3.8.6	Smart Redundancy mode	59
3.9	Specifications	59
3.9.1	Physical specifications	59
3.9.2	Supported environment.	60
Chapter 4.	Compute nodes	61
4.1	NeXtScale nx360 M5 compute node	62
4.1.1	Overview	63
4.1.2	System architecture	67
4.1.3	Standard specifications	69
4.1.4	Standard models	72
4.1.5	Processor options	73
4.1.6	Memory options	74
4.1.7	NeXtScale 12G Storage Native Expansion Tray	79
4.1.8	Internal storage	80
4.1.9	Controllers for internal storage	83
4.1.10	Internal drive options	88
4.1.11	I/O expansion options	92
4.1.12	Network adapters	93
4.1.13	Storage host bus adapters	95
4.1.14	NeXtScale PCIe Native Expansion Tray	96
4.1.15	NeXtScale PCIe 2U Native Expansion Tray	98
4.1.16	GPU and coprocessor adapters	100
4.1.17	Integrated virtualization	102
4.1.18	Local server management	103
4.1.19	Remote server management	104
4.1.20	Supported operating systems	106
4.1.21	Physical and environmental specifications	107
4.1.22	Regulatory compliance	108
4.2	NeXtScale nx360 M4 compute node	110
4.2.1	Overview	110
4.2.2	System architecture	113
4.2.3	Specifications	115
4.2.4	Standard models	117
4.2.5	Processor options	117
4.2.6	Memory options	118
4.2.7	Internal disk storage options	123
4.2.8	NeXtScale Storage Native Expansion Tray	130
4.2.9	NeXtScale PCIe Native Expansion Tray	133
4.2.10	GPU and coprocessor adapters	134
4.2.11	Embedded 1 Gb Ethernet controller	137
4.2.12	PCI Express I/O adapters	138
4.2.13	Integrated virtualization	142
4.2.14	Local server management	143
4.2.15	Remote server management	144
4.2.16	External disk storage expansion	146
4.2.17	Physical specifications	148

4.2.18 Operating systems support	149
Chapter 5. Rack planning	151
5.1 Power planning	152
5.1.1 NeXtScale Rack Power Reference Examples	153
5.1.2 Examples	154
5.1.3 PDUs.	156
5.1.4 UPS units	157
5.2 Cooling	157
5.2.1 Planning for air cooling	157
5.3 Density	159
5.4 Racks	159
5.4.1 Rack Weight	159
5.4.2 The 42U 1100mm Enterprise V2 Dynamic Rack	160
5.4.3 Installing NeXtScale System in other racks	165
5.4.4 Shipping the chassis.	166
5.4.5 Rack options	167
5.5 Cable management.	171
5.6 Rear Door Heat eXchanger.	173
5.7 Top-of-rack switches	175
5.7.1 Ethernet switches	176
5.7.2 InfiniBand switches	176
5.7.3 Fibre Channel switches.	177
5.8 Rack-level networking: Sample configurations	177
5.8.1 Non-blocking InfiniBand	178
5.8.2 A 50% blocking InfiniBand	179
5.8.3 10 Gb Ethernet, one port per node	180
5.8.4 10 Gb Ethernet, two ports per node	181
5.8.5 Management network	182
Chapter 6. Factory integration and testing	183
6.1 Lenovo Intelligent Cluster	184
6.2 Lenovo factory integration standards	184
6.3 Factory testing.	185
6.4 Documentation provided	187
6.4.1 HPLinpack testing results: Supplied on request	187
Chapter 7. Managing a NeXtScale environment.	189
7.1 Managing compute nodes.	190
7.1.1 Integrated Management Module II	190
7.1.2 Unified Extendible Firmware Interface	193
7.1.3 ASU.	203
7.1.4 Firmware upgrade.	204
7.2 Managing the chassis	205
7.2.1 FPC web browser interface.	205
7.2.2 FPC IPMI interface	224
7.3 ServeRAID C100 drivers: nx360 M4	232
7.4 Integrated SATA controller: nx360 M5	232
7.5 VMware vSphere Hypervisor	232
7.6 eXtreme Cloud Administration Toolkit.	233
Abbreviations and acronyms	237
Related publications	241

Lenovo Press publications 241

Product publications 241

Online resources 242

Notices

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

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

Lenovo (United States), Inc.
1009 Think Place - Building One
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing

LENOVO 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 jurisdictions 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. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary.

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

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

Any performance data contained herein was determined in a controlled environment. Therefore, the result 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.

Trademarks

Lenovo, the Lenovo logo, and For Those Who Do are trademarks or registered trademarks of Lenovo in the United States, other countries, or both. These and other Lenovo 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 Lenovo at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of Lenovo trademarks is available on the Web at <http://www.lenovo.com/legal/copytrade.html>.

The following terms are trademarks of Lenovo in the United States, other countries, or both:

Advanced Settings Utility™	Lenovo®	ServerGuide™
BladeCenter®	NeXtScale™	ServerProven®
Dynamic System Analysis™	NeXtScale System®	System x®
Flex System™	RackSwitch™	ThinkServer®
iDataPlex®	Lenovo(logo)®	TruDDR4™
Intelligent Cluster™	ServeRAID™	UpdateXpress System Packs™

The following terms are trademarks of other companies:

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, Windows, 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

Lenovo® NeXtScale™ System is a dense computing offering that is based on our experience with iDataPlex® and Flex System™ and with a tight focus on emerging and future client requirements. The NeXtScale n1200 Enclosure and NeXtScale nx360 M5 Compute Node optimize density and performance within typical data center infrastructure limits.

The 6U NeXtScale n1200 Enclosure fits in a standard 19-inch rack and up to 12 compute nodes can be installed into the enclosure. With more computing power per watt and the latest Intel Xeon processors, you can reduce costs while maintaining speed and availability.

This Lenovo Press publication is for customers who want to understand and implement a NeXtScale System® solution. It introduces the offering and the innovations in its design, outlines its benefits, and positions it with other x86 servers. The book provides details about NeXtScale System components and the supported options. It also provides rack and power planning considerations and describes the ways that you can manage the system.

This book describes the air-cooled NeXtScale System offerings. For more information about planning and implementation of the water-cooled offering, NeXtScale System WCT, see *Lenovo NeXtScale System Water Cool Technology Planning and Implementation Guide*, SG24-8276, which is available at this website:

<http://lenovopress.com/sg248276>

Lenovo NeXtScale System servers are based on Intel Xeon processors.



The team who wrote this book

This document is produced by the following subject matter experts working in the Lenovo offices in Morrisville, NC, US:

David Watts is a Senior IT Consultant and the program lead for Lenovo Press. He manages residencies and produces pre-sale and post-sale technical publications for hardware and software topics that are related to System x, ThinkServer, Flex System, and BladeCenter servers. He has authored over 300 books and papers. David has worked in the IT industry, both in the U.S. and Australia, since 1989, and is currently based in Morrisville, North Carolina. David holds a Bachelor of Engineering degree from the University of Queensland (Australia).

Matthew Archibald is the Global Installation and Planning Architect for Lenovo Professional Services. He has been with System x for 12 years and worked for the last nine years in the data center space. Before working in his current position, Matt worked as a development engineer in the System x and BladeCenter power development lab and was responsible for power subsystem design for BladeCenter and System x servers. Matt is also responsible for the development, maintenance, and support of the System x Power Configurator program and holds 29 patents for various data center and hypervisor technologies. Matt has degrees from Clarkson University in Computer Engineering, Electrical Engineering, Software Engineering, and Computer Science and a Bachelor of Engineering in Electronics Engineering from Auckland University of Technology.

Jerrold Buterbaugh is the Worldwide Data Center Services Principal for Lenovo Professional Services. He has been a Data Center practice lead for the last seven years, focused on facility and IT installations and optimization. Previously, Jerrold worked as a power development engineer in the System x and IBM Power Systems power development labs designing server power subsystems. Jerrold is also the key focal point in supporting HPC installations and holds 18 patents for various data center and server technologies.

Duncan Furniss is a Consulting Client Technical Specialist for Lenovo in Canada. He provides technical sales support for iDataPlex, NeXtScale, BladeCenter, Flex, and System x products. He co-authored several Lenovo Press and IBM Redbooks publications, including *NeXtScale System M5 with Water Cool Technology* and *NeXtScale System Planning and Implementation Guide*. Duncan also designed and provided oversight for the implementation of many large-scale solutions for HPC, distributed databases, and rendering of computer-generated images.

David Latino is the lead HPC Architect and HPC Center of Competency leader for Lenovo Middle East and Africa. Before working for Lenovo starting in 2015, he was a Consulting IT Specialist, performing the same role for IBM Middle East, Turkey, and Africa. David has 12 years of experience in the HPC field. He led a wide spectrum of consulting projects, working with HPC users in academic research and industry sectors. His work covered many aspects of the HPC arena and he was technical leader for the design and implementation of multiple large HPC systems that appeared in the top500 list. David worked extensively on HPC application development, optimization, scaling, and performance benchmark evaluation, which resulted in several highly optimized application software packages. He also spent several years based at customer sites to train system administrators, users, and developers to manage and efficiently use IBM Blue Gene systems.

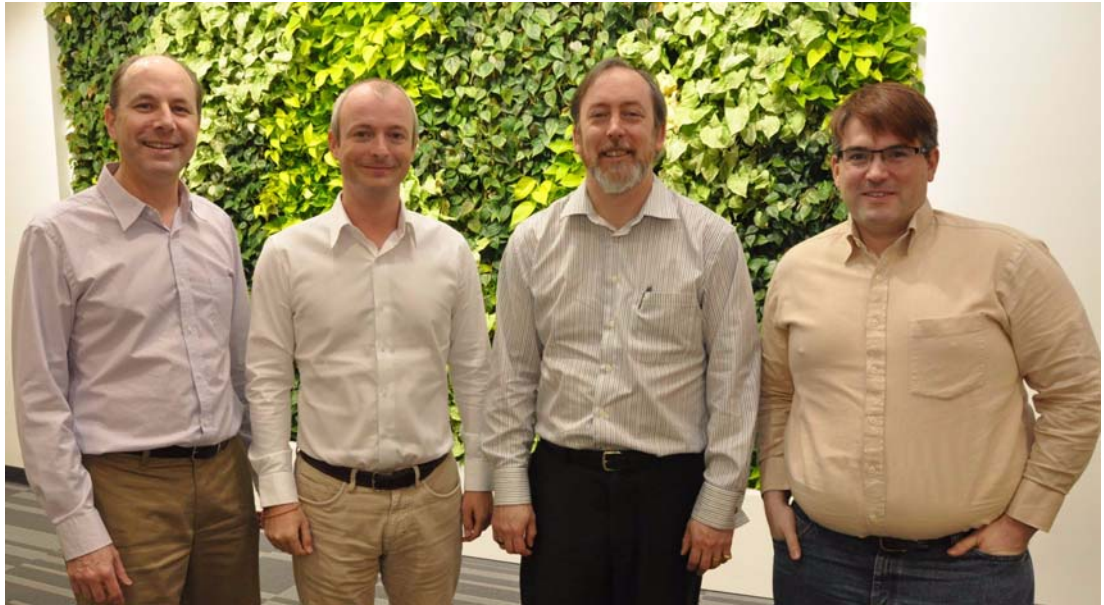


Figure 1 Four of the team (l-r): David Watts, David Latino, Duncan Furniss, and Matt Archibald

Thanks to the following authors of the previous editions of this book:

- ▶ David Watts
- ▶ Jordi Caubet
- ▶ Duncan Furniss
- ▶ David Latino

Thanks to the following people who contributed to the project:

Lenovo Press: Ilya Krutov

System x Marketing:

- ▶ Jill Caugherty
- ▶ Keith Taylor
- ▶ Scott Tease

NeXtScale Development:

- ▶ Vinod Kamath
- ▶ Edward Kung
- ▶ Mike Miller
- ▶ Wilson Soo
- ▶ Mark Steinke

Comments welcome

Your comments are important to us!

We want our books to be as helpful as possible. Send us your comments about this book by using one of the following methods:

- ▶ Use the online **Contact us** review Redbooks form that is found at:
ibm.com/redbooks
- ▶ Send your comments in an email to:
redbooks@us.ibm.com

Summary of changes

This section describes the technical changes that were made in this edition of the Interoperability Guide and in previous editions. This edition might also include minor corrections and editorial changes that are not identified.

August 2015

This revision reflects the addition, deletion, or modification of new and changed information that is described in this section.

New products

- ▶ NeXtScale nx360 M5
- ▶ New Intel Xeon E5-2600 v3 processor options
- ▶ New DIMM options
- ▶ New drive options
- ▶ NeXtScale 12G Storage Native Expansion Tray
- ▶ NeXtScale PCIe 2U Native Expansion Tray
- ▶ n1200 Enclosure Fan and Power Controller 2
- ▶ CFF -48V DC power supply option
- ▶ 1500 W AC power supply option
- ▶ 1300 W AC Titanium power supply option
- ▶ UPS units
- ▶ RackSwitch™ options

Revised information

- ▶ Power supply tables
- ▶ Cabling recommendations
- ▶ Rack weight information

July 2014

This revision reflects the addition, deletion, or modification of new and changed information that is described in this section.

New information

- ▶ Added 1300W power supply efficiency values, Table 3-31 on page 46
- ▶ Added tables showing quantities of compute nodes that are supported based on processor selection, power supply selection, and input voltage, 3.3, “Supported compute nodes” on page 25
- ▶ Added information about GPUs, 4.2.10, “GPU and coprocessor adapters” on page 134

April 2014, Second Edition

This revision reflects the addition, deletion, or modification of new and changed information that is described in this section.

New information

- ▶ New PCIe Native Expansion Tray supporting GPUs and coprocessors
- ▶ New Intel Xeon Phi coprocessor and NVIDIA GPU adapter options
- ▶ New Intel Xeon E5-2600 v2 processor options
- ▶ New 1300W power supply option
- ▶ New models of the NeXtScale n1200 chassis with 1300 W power supplies
- ▶ New RDIMM memory options
- ▶ Support for 2.5-inch SSD options and other drive options
- ▶ Support for ServeRAID™ M5120 RAID controller for external SAS storage expansion

Introduction

NeXtScale System is the next generation of dense computing. It is an open, flexible, and simple data center solution for users of technical computing, grid deployments, analytics workloads, and large-scale cloud and virtualization infrastructures.

NeXtScale System is built with industry-standard components to create flexible configurations of servers, chassis, and networking switches that integrate easily in a standard 19-inch rack. It is a general-purpose platform that provides flexibility to clients for creating unique and differentiated solutions by using off-the-shelf components. Front-access cabling enables you to quickly and easily make changes in networking and power connections.

The NeXtScale n1200 Enclosure and NeXtScale nx360 M5 server are the major components of the offering. These components optimize density and performance within typical data center infrastructure limits. The 6U n1200 enclosure fits in a standard 19-inch rack and up to 12 nx360 M5 servers can be installed into the enclosure.

The nx360 M5 and n1200 enclosure are also available in direct-water cooled configurations for the ultimate in data center cooling efficiencies, as described in *NeXtScale System Water Cooled Planning and Implementation Guide*, SG24-8276, which is available at this website:

<http://lenovopress.com/sg248276>

This chapter includes the following topics:

- ▶ 1.1, “Evolution of the data center” on page 2
- ▶ 1.2, “Summary of the key components” on page 3
- ▶ 1.3, “Design points of the system” on page 6
- ▶ 1.5, “This book” on page 7

1.1 Evolution of the data center

There is an increasing number of computational workloads that can be run on groups of servers, which are often referred to by such names as clusters, farms, or pools. This type of computing can be described as scale-out; however, as a convention, we refer to these groups as *clusters*. As the computing community's proficiency with implementing and managing clusters improved, there is a trend to create large clusters, which are becoming known as *hyper-scale environments*.

In the past, when the number of servers in a computing environment was lower, considerable hardware engineering effort and server cost was expended to create servers that were highly reliable to reduce application downtime. With clusters of servers, we strive to create a balance between the high availability technologies that are built in to every server and reduce the cost and complexity of the servers, which allows more of them to be provisioned.

The mainstream adoption of virtualization and cloud software technologies in the data center caused a paradigm shift at the server level that further removes the reliance on high availability hardware. The focus is now on providing high availability applications to users on commodity hardware with workloads that are managed and brokered with highly recoverable virtualization platforms. With this shift away from hardware reliability, new server deployment methods emerged that allow previous data center barriers to grow without the need for large capital investments.

1.1.1 Density

As the number of servers in clusters grows and the cost of data center space increases, the number of servers in a unit of space (also known as the *compute density*) becomes an increasingly important consideration. NeXtScale System optimizes density while addressing other objectives, such as, providing the best performing processors, minimizing the amount of energy that is used to cool the servers, and providing a broad range of configuration options.

Increased density brings new challenges for facility managers to cool high-performance, highly dense rack-level solutions. To support this increased heat flux, data center facilities teams are investigating the use of liquid cooling at the rack. NeXtScale System was designed with this idea in mind, in that it can use traditional air cooling or can be cooled by using the latest Water Cool Technology.

1.1.2 Scale out applications

The following applications are among the applications that lend to clusters of servers:

- High performance computing (HPC)

HPC is a general category of applications that are computationally complex, can deal with large data sets, or consist of vast numbers of programs that must be run. Examples of computationally complex workloads include weather modeling or simulating chemical reactions. Comparing gene sequences is an example of a workload that involves large data sets. Image rendering for animated movies and Monte Carlo analysis for particle physics are examples of workloads where there are vast numbers of programs that must be run. The use of several HPC clusters in a Grid architecture is an approach that gained popularity.

- Cloud services

Cloud services that are privately owned and those services that are publicly available from managed service providers provide standardized computing resources from pools of homogeneous servers. If a consumer requires more or less server capacity, the servers are provisioned from or returned to the pools. This paradigm often also includes consumer self-service and usage metering with some form of show back, charge back, or billing.

- Analytics

Distributed databases and extensive use of data mining, or analytics, is another use case that is increasing in prevalence and is applied to a greater range of business and technical challenges.

1.2 Summary of the key components

The NeXtScale n1200 Enclosure and NeXtScale nx360 M5 server optimize density and performance within typical data center infrastructure limits. The 6U n1200 enclosure fits in a standard 19-inch rack and up to 12 nx360 M5 servers can be installed into the enclosure.

NeXtScale System is based on a simple chassis configuration with shared power and cooling. The chassis is a standard 19-inch rack component that makes it easy to integrate with your networking, storage, and data center infrastructure components.

The NeXtScale nx360 M5 server provides a dense, flexible solution with a low total cost of ownership (TCO). The half-wide, dual-socket NeXtScale nx360 M5 server is designed for data centers that require high performance but are constrained by floor space. By taking up less physical space in the data center, the NeXtScale server enhances density and it supports the Intel Xeon processor E5-2600 v3 series up to 145 W and 18-core processors, which provides more performance per server. The nx360 M5 compute node contains only essential components in the base architecture to provide a cost-optimized platform.

The nx360 M5 also supports more expansion options in the form of trays that attach to the top of the server. The PCIe Native Expansion Tray can be added to the nx360 M5 to form a powerful compute engine that supports two GPU or coprocessor adapters. The Storage Native Expansion Tray can be added to the nx360 M5 to form a storage-dense server that supports up to 48 TB of local storage.

The NeXtScale n1200 Enclosure is an efficient, 6U, 12-node chassis with no built-in networking or switching capabilities; therefore, it requires no chassis-level management. Sensibly designed to provide shared, high-efficiency power and cooling for housed servers, the n1200 enclosure scales with your business needs.

The NeXtScale nx360 M5 also is available as a warm-water-cooled server for the ultimate in energy efficiency, cooling, noise, and TCO. For more information, *NeXtScale System Water Cooled Planning and Implementation Guide*, SG24-8276, which is available at this website:

<http://lenovopress.com/sg248276>

1.2.1 Lenovo NeXtScale n1200 Enclosure

The NeXtScale System is based on a six-rack unit (6U) high chassis with 12 half-width bays, as shown in Figure 1-1.



Figure 1-1 Front of NeXtScale n1200 enclosure, with 12 half-wide compute nodes

Chassis power and cooling

The NeXtScale n1200 Enclosure includes 10 hot swappable fans and six hot swappable power supplies, which are installed in the rear of the chassis, as shown in Figure 1-2.



Figure 1-2 Rear of NeXtScale n1200 Enclosure with 900 W power supplies shown

Six single-phase power supplies enable power feeds from one or two sources of three-phase power.

Also, in the rear of the chassis is the Fan and Power Controller, which controls power and cooling aspects of the chassis.

1.2.2 NeXtScale compute nodes

NeXtScale System addresses today's computing challenges by offering two generations of compute nodes. Both compute nodes can be configured with native expansion trays to more efficiently address compute, storage, I/O, and acceleration workloads.

NeXtScale nx360 M5 Compute Node

The newest server available for the NeXtScale System is the nx360 M5 Compute Node. As with its predecessor, the nx360 M5 fits in a half-wide bay in the n1200 Enclosure, as shown in Figure 1-1 on page 4.

The server is shown in Figure 1-3. On the front is the power button, status LEDs, and connectors. There is a full-height, half-length PCI Express card slot and a PCI Express mezzanine card slot that uses the same mezzanine card type as our rack mount servers.

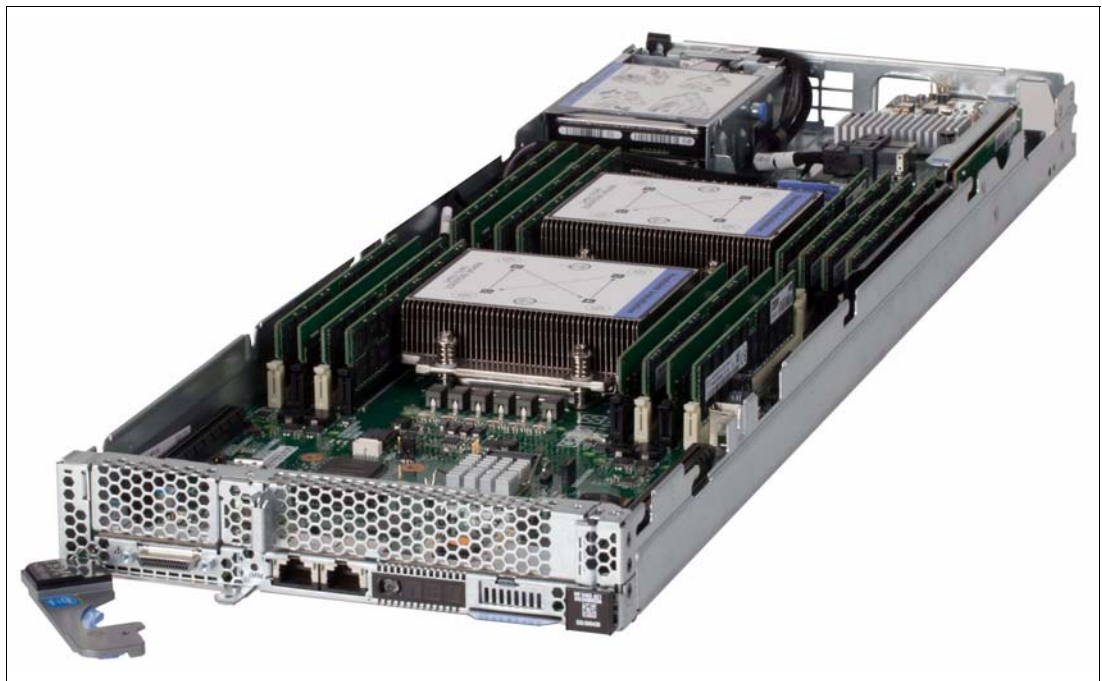


Figure 1-3 NeXtScale nx360 M5

Inside, the nx360 M5 supports two Intel Xeon E5-2600 v3 series processors, 16 DDR4 DIMMs, and a hard disk drive (HDD) carrier. HDD carrier options include one 3.5-inch drive, two 2.5-inch drives, or four 1.8-inch solid-state drives (SSD).

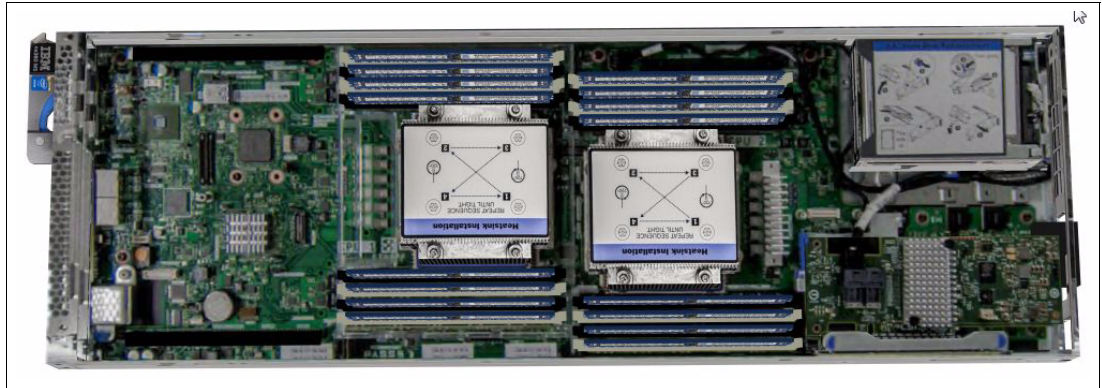


Figure 1-4 IBM NeXtScale nx360 M5 with two 2.5-inch HDDs

NeXtScale nx360 M4 compute node

The first server that is available for the NeXtScale System is the nx360 M4 compute node. It fits in a half-width bay in the n1200 Enclosure, as shown in Figure 1-5. On the front is the power button, status LEDs, and connectors. There is a full-height, half-length PCI Express card slot, and a PCI Express mezzanine card slot that uses the same mezzanine card type as our rack mount servers.

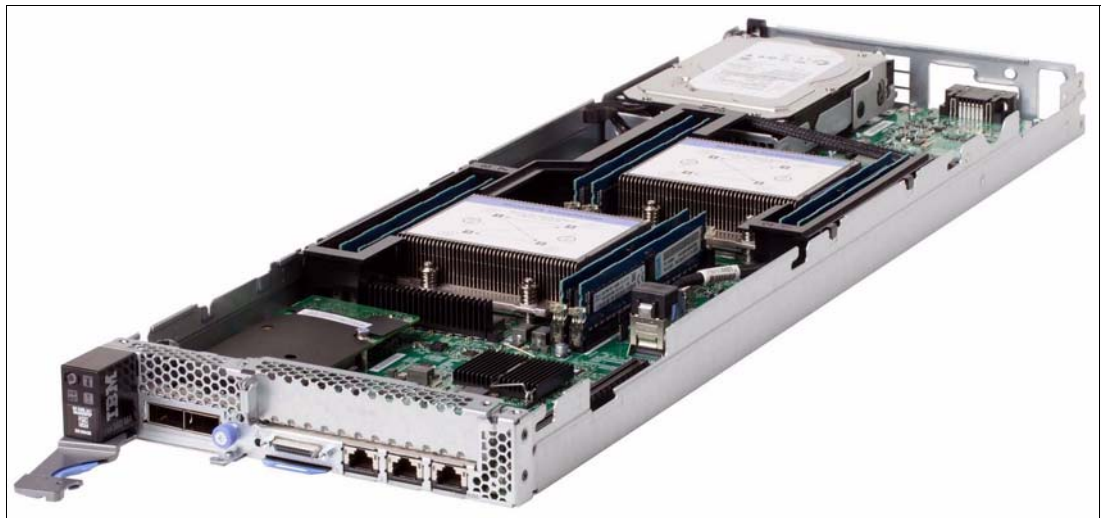


Figure 1-5 NeXtScale nx360 M4

Inside, the nx360 M4 supports two Intel Xeon E5-2600 v2 series processors, eight DDR3 DIMMs, and an HDD carrier. HDD carrier options include one 3.5-inch drive, two 2.5-inch drives, or four 1.8-inch SSDs. The server is shown in Figure 1-6.

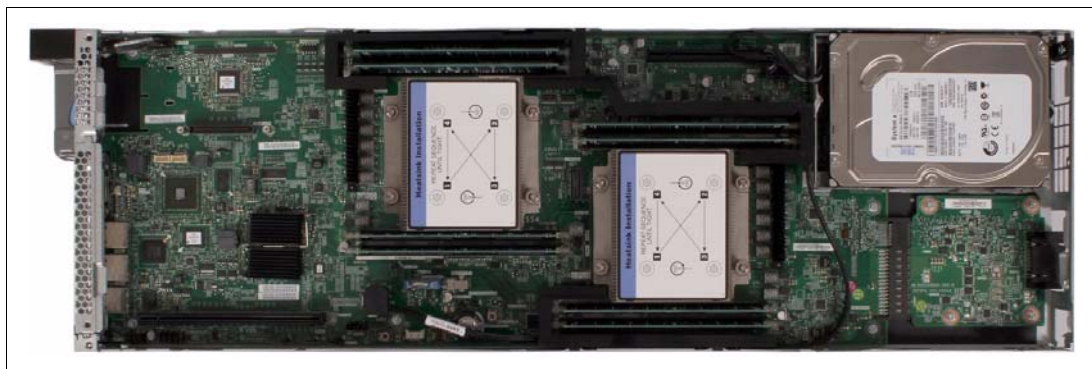


Figure 1-6 NeXtScale nx360 M4 with one 3.5-inch hard disk drive

1.3 Design points of the system

This section describes some of the following design points that are included in the Lenovo NeXtScale System:

- System is designed for flexibility

The power supplies and fans in the back of the chassis are modular and hot swappable. The servers slide in and out of the front and have their cable connections at the front.

- The chassis supports devices that span multiple bays.

Compute node designs are not limited to a single 1U-high half-wide server. As with iDataPlex, servers can be augmented with trays that enable more features, such as, adapters or drives. More systems that extend the design are in development.

- Fits in a standard rack

The NeXtScale n1200 Enclosure can be installed into many standard 19-inch racks (which might require more cable routing brackets) and the rack can have a mixture of NeXtScale and other components. A system can start with a few servers and grow incrementally. Alternatively, you can have Lenovo install the servers into racks with switches and power distribution units and all of the cables connected.

- Factory integration available

More configuration and testing are done when the systems are factory-integrated. For more information about Lenovo factory integration, see Chapter 6, "Factory integration and testing" on page 101.

- NeXtScale System is focused on computational density

Compared to an iDataPlex system with 84 servers in each iDataPlex rack, with six NeXtScale chassis in each 42U standard rack (which leaves 6U per rack for more components), 28% more servers can be fit in the same floor tile configuration. With standard racks, clients can design compact data center floor layouts for all their equipment.

- System is designed for simplicity

Cable access to the server is from the front. The servers are directly accessed for their local console, management, and data networks, which eliminates contention.

- Uses standard components

The servers support standard PCI Express (Generation 3) adapters and have RJ-45 copper Ethernet interfaces on board. The chipsets that are used in the server were selected with broad industry acceptance in mind.

1.4 NeXtScale System cooling choices

NeXtScale System is a dense IT solution that is deployable in various data center cooling environments from traditional forced air, raised floor data centers to highly efficient data centers that distribute water to the IT racks and use free cooling.

The following cooling solutions are supported by NeXtScale System:

- Air-cooled

A traditional cooling topology in which the cooled air is delivered to the IT systems at the front of the rack and the hot exhaust air exits the rear of the rack and is returned to the forced air cooling system.

- Air-cooled and water-cooled with Rear Door Heat Exchanger (RDHX)

Similar to the traditional cooling topology in which cool air is delivered to the front of the rack, but the heat that exits the IT systems is immediately captured by the water-cooled RDHX and returned to the facility chilled water cooling system.

- Direct water-cooled (WCT)

A non-traditional cooling approach in which cool or warm water is distributed directly to the system CPUs, memory, and other subsystem heat sinks via a rack water manifold assembly.

For more information about the water-cooled solutions, see Lenovo Press book *Lenovo NeXtScale System Water Cool Technology Planning and Implementation Guide*, SG24-8276, which is available at this website:

<http://lenovopress.com/sg248276>

- Hybrid - direct water-cooled (WCT) coupled with RDHX

A non-traditional cooling approach that combines WCT with an external RDHX to remove 100% of the rack heat load and provides a room neutral solution.

For more information about the benefits of the use of an RDHX with NeXtScale WCT, see *Lenovo NeXtScale System Water Cool Technology Planning and Implementation Guide*, SG24-8276, which is available at this website:

<http://lenovopress.com/sg248276>

1.5 This book

In this book, we compare NeXtScale System to other systems and raise points to help you select the right systems for your applications. We then take an in-depth look at the chassis, servers, and fan and power controller (FPC). Next, we take a broader view and cover implementations at scale and review racks and cooling. We then describe Lenovo's process for assembling and testing complete systems in to Intelligent Cluster™ solutions.

Positioning

NeXtScale is ideal for fastest-growing workloads, such as, social media, analytics, technical computing, and cloud delivery, which are putting increased demands on data centers.

This chapter describes how NeXtScale System is positioned in the marketplace compared with other systems that are equipped with Intel processors. The information helps you to understand the NeXtScale target audience and the types of workloads for which it is intended.

This chapter includes the following topics:

- ▶ 2.1, “Market positioning” on page 10
- ▶ 2.2, “System x and ThinkServer overview” on page 14
- ▶ 2.3, “NeXtScale System versus iDataPlex” on page 15
- ▶ 2.4, “NeXtScale System versus Flex System” on page 16
- ▶ 2.5, “NeXtScale System versus rack-mounted servers” on page 18

2.1 Market positioning

NeXtScale System is a new x86 offering that introduces a new category of dense computing into the marketplace. NeXtScale System includes the following key characteristics:

- ▶ Strategically, this system is the next generation dense system from Lenovo that includes the following features:
 - A building block design that is based on a low function and low-cost chassis.
 - Flexible compute node configurations that are based around a 1U half-wide compute node supports various application workloads.
 - A standard rack platform.
- ▶ Built for workloads that require density
- ▶ NeXtScale performs well in scale-out applications, such as, cloud, HPC, grid, and analytics
- ▶ Is central in OpenStack initiatives for public clouds

NeXtScale System includes the following key features:

- ▶ Supports up to seven chassis¹ in a 42U rack, which means up to a total of 84 systems and 3,024 processor cores in a standard 19-inch rack.
- ▶ Industry-standard components for flexibility, ease of maintenance, and adoption.
- ▶ Approved for data centers with up to 40°C ambient air temperature, which lowers cooling costs.
- ▶ Available as single node, an empty or configured chassis, or in full racks.
- ▶ Can be configured as part of the Intelligent Cluster processor for complete pre-testing, configuration, and arrival ready to plug in.
- ▶ Compute nodes offer the fastest Intel Xeon processors (top-bin 145 W) with new 2133 MHz memory.
- ▶ Supports 100 - 127 V and 200 - 240 V power.
- ▶ Standard form factor and components make it ideal for Business Partners.

Direct water cooling includes the following benefits:

The customer that benefits the most from NeXtScale is an enterprise that is looking for a low-cost, high-performance computing system to start or optimize cloud, big data, Internet, and technical computing applications, which include the following uses:

- ▶ Large data centers that require efficiency, density, scale, and scalability.
- ▶ Public, private, and hybrid cloud infrastructures.
- ▶ Data analytics applications, such as, customer relationship management, operational optimization, risk and financial management, and enabling new business models.
- ▶ Internet media applications, such as, online gaming and video streaming.
- ▶ High-resolution imaging for applications ranging from medicine to oil and gas exploration.
- ▶ “Departmental” uses in which a small solution can increase the speed of outcome prediction, engineering analysis, and design and modeling.

¹ Six chassis per rack are recommended because this amount leaves rack space for switches and cable routing. The use of seven chassis in a rack might require removal of the rack doors.

2.1.1 Three key messages with NeXtScale

The three key messages about NeXtScale System is that it is flexible, simple, and scalable, as shown in Figure 2-1.

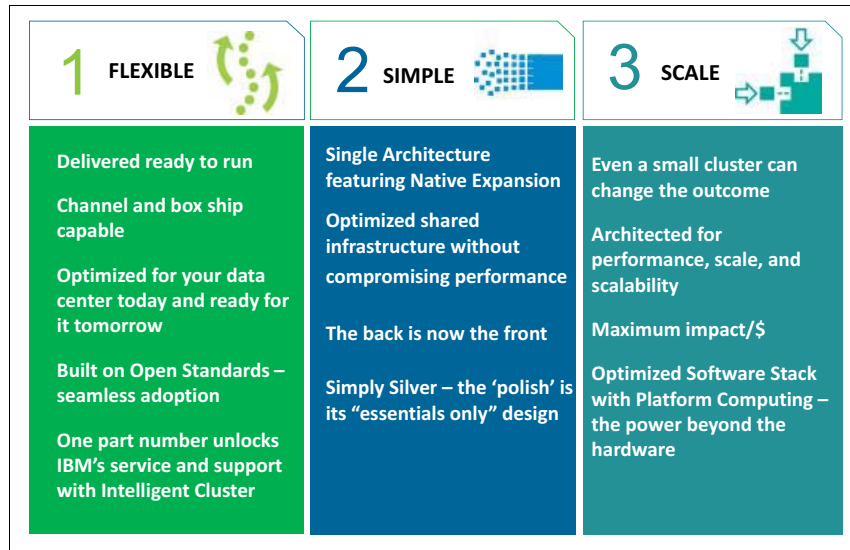


Figure 2-1 NeXtScale System key messages

NeXtScale System is flexible in the following ways:

► Ordering and delivery

The question of how you want your NeXtScale System configuration to be ordered and delivered is complex because there are many choices. Some clients want everything in parts so that they can mix and match to build what they want. Others want systems that they tested and approved to show that it is configured to their liking. Still others want complete solutions of racks to arrive ready to plug in. With NeXtScale, the choice is yours.

► A hardware design that allows for a mix of compute nodes.

The NeXtScale compute nodes are 1U half-wide servers but are extended with the addition of various trays (also known as *Native Expansion*) with which you can select the systems that you need that is based on the needs of the applications that you run.

► Fit it into your data center seamlessly in a Lenovo rack or most 19-inch standard racks.

The NeXtScale n1200 Enclosure is installed in the 42U 1100mm Enterprise V2 Dynamic Rack because it provides the best cabling features. However, the chassis can also be installed in many third-party, 4-post, 19-inch racks. This flexibility ensures maximum flexibility regarding deploying NeXtScale System into your data center.

► NeXtScale System is backed by leading Lenovo service and support no matter how you buy it, where you use it, or what task you have it running.

► Support for open standards.

A client needs more than hardware to use IT. We designed NeXtScale to support an open stack of industry standard tools to allow clients that have protocols and tools to migrate easily by using NeXtScale System.

The nx360 M4 and nx360 M5 compute nodes offers the Integrated Management Module II service processor and the n1200 Enclosure has the Fan and Power Controller. Both support the IPMI protocol for flexible and standards-based systems managements.

NeXtScale System is simple in the following ways:

- ▶ A design that is based on the half-wide compute node.

The architecture of NeXtScale System revolves around a low-function chassis that hosts compute nodes. The design supports Native Expansion that allows seamless upgrades to add common functionality, such as, storage, graphics acceleration, or co-processing at the time of shipment or in the future.

- ▶ A chassis that includes shared fans and cooling.

The n1200 Enclosure supplies the cooling and power to maximize energy efficiency, but leaves management and connectivity to the compute nodes, which minimizes cost.

- ▶ Cables, connectors, and controls at the front.

Except for the power cords, all cabling is at the front of the chassis. All controls and indicators also are at the front. This configuration makes access, server swap-outs, and overall systems management easier.

Each compute node has a front connector for a local console for use if you need crash-cart access. Also, each compute node has a pull-out tab at the front for system and customer labeling needs, as shown in Figure 2-2.

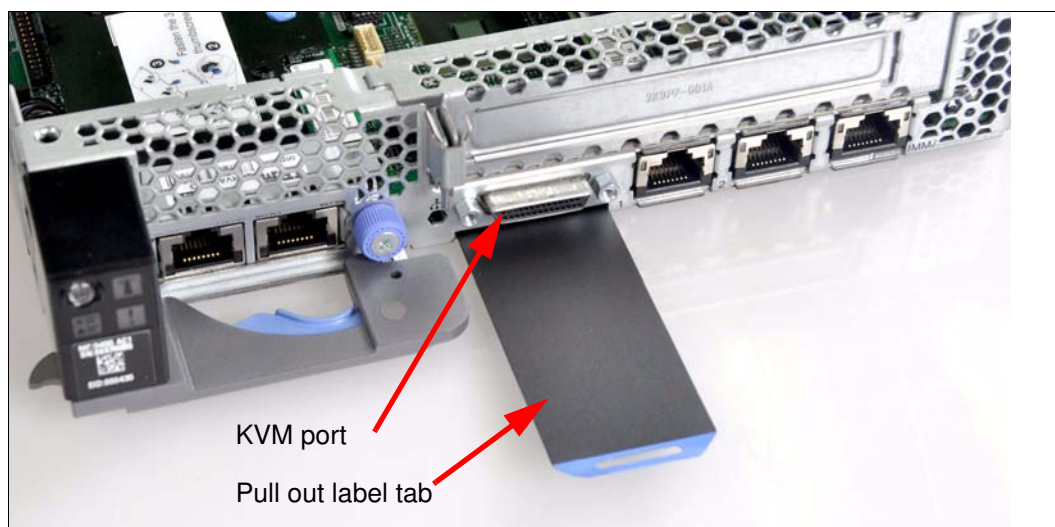


Figure 2-2 Front of the NeXtScale nx360 M4

Because the cables do not clog up the back of the rack, air flow is improved and thus energy efficiency also is improved. The harder the fans work to move air through server, the more power they use.

Your support staff who work in the data center can tell you the front of the rack is a much more enjoyable environment to spend time in because it might easily be 30 °F (16 °C) cooler at the front than at the back. People cannot stay in the rear of the rack for long before it is no longer comfortable. Also, the front of rack is less noisy than the rear of the rack because of fan noise.

It is also difficult to locate a single dense server in a row of dense racks and then go to the back to service the cabling. Having all of the cabling on the front simplifies and reduces the chances of mis-cabling or pulling the wrong server.

- ▶ Installation in a three-phase power data center.

The design of six power supplies per chassis allows seamless installation into data centers with three-phase power. With six supplies and two, three-phase feeds, power delivery is optimized and balanced; there is no waste, no inefficiency.

- ▶ The compute nodes are unpainted metal.
- ▶ Unlike every other x86 server Lenovo offers, these servers do not have a black front to them, which indicates simplicity and efficiency.

NeXtScale System is scalable in the following ways:

- ▶ Scaling is for everyone

As we describe scale, it is important to understand that scale is not for massive deployments only; even a small, one-chassis solution can change what users believe can be done.

Whether you start small and grow or start huge and grow enormously, NeXtScale can be run and managed at scale as a single solution.

- ▶ NeXtScale System is built on what was learned about the financial aspects of scale-out.

Every decision about the product was aimed at improving our clients impact per dollar, whether that meant removing components that are not required or by selecting more energy efficient parts to reduce power usage and, therefore, power costs.

- ▶ Scalable to the container level.

NeXtScale System can meet the needs of clients who want to add IT at the rack level or even at the container level. Racks can be fully configured, cabled, labeled, and programmed before they are shipped. Lenovo also can take configured racks and assemble them into complete, containerized solutions with power, cooling, and infrastructure delivered ready to go.

2.1.2 Optimized for workloads

NeXtScale System is best-suited for the following primary workloads:

- ▶ Public and private cloud
- ▶ HPC and technical computing

Although these areas do have much in common, they also have unique needs that are served with NeXtScale System, as shown in Figure 2-2.

Workload	Fine Tuned Server Characteristics
<div>Private Cloud</div> <div>Public Cloud</div>	<ul style="list-style-type: none"> • Processor and Memory performance and choice Full Intel stack support with memory for performance and/or cost optimization • Standard Rack optimized Fits into client data centers seamlessly • Right sized IO Choice of networking options – 1Gb, 10Gb, or InfiniBand, all SDN ready • Infinitely Scalable from small to enormous grid deployments all built on open standards • High energy efficiency means more impact/watt
<div>High Performance Computing</div>	<ul style="list-style-type: none"> • Top bin Intel Xeon processors, large memory bandwidth, and high IOPS for rapid transaction processing and analytics • Workload optimized software stack with Platform Computing and IBM xCAT • Architected for low latency with choice of high speed fabric support • Supported as one part number no matter the size of the solution and content with Intelligent Cluster

Figure 2-3 NeXtScale System: Optimized for cloud computing and HPC solutions

For cloud, the important factors are that the entire processor stack is supported; therefore, no matter what the client’s goal is, it can be supported by the right processor performance and cost point. The same is true of memory; cost and power-optimized choices and performance-optimized alternatives are available. The nodes feature the networking on board with 1 Gb NICs embedded, with options for up to four other high-speed fabric ports. With these features, the entire solution can scale to any size.

HPC and technical computing have many of the same attributes as cloud; a key factor is the need for the top-bin 145 W processors. NeXtScale System can support top bin 145 W processors, which means more cores and higher frequencies than others.

2.2 System x and ThinkServer overview

The world is evolving, and the way that our clients do business is evolving with it. That is why Lenovo has the broadest x86 portfolio in our history and is expanding even further to meet the needs of our clients, whatever those needs might be.

The x86 server market is segmented on the following key product areas:

- High-end systems

Lenovo dominates this space with enterprise-class X6 four-socket and eight-socket server that offer unprecedented x86 performance, resiliency, and security.
- Blades and integrated systems

Integrated systems is a fast growing market where Lenovo adds value by packaging Lenovo software and hardware assets in a way that helps our clients optimize value.

- Dense systems

As with NeXtScale or iDataPlex, dense systems is a fast growing segment that is pushed by data center limitations and new workloads that require scale out architecture. These systems transformed how clients optimize space-constrained data centers with extreme performance and energy efficiency.

- High volume systems

The volume space is over half the total x86 server market and Lenovo has a broad portfolio of rack and tower servers to meet a wide range of client needs, from infrastructure to technical computing. This portfolio includes offerings from the System x and ThinkServer® brands.

- System Networking

System Networking solutions are designed for complex workloads that are tuned for performance, virtualization, and massive scale. Lenovo takes an interoperable, standards-based approach to implement the latest advances in today's high-speed, converged data center network designs with optimized applications and integrated systems.

- Enterprise Storage

Lenovo delivers simplified, centralized storage solutions for small businesses to enterprises with excellent performance and reliability, advanced data protection, and virtualization capabilities for your business-critical data.

2.3 NeXtScale System versus iDataPlex

Although iDataPlex and NeXtScale look different, many of the ideas and innovations we pioneered with iDataPlex remain in the new NeXtScale System.

When we introduced iDataPlex in 2008, we introduced a chassis that was dedicated to power and cool independent nodes. With NeXtScale system, we reuse the same principle, but we are extending it to bring more flexibility to the users.

The NeXtScale n1200 Enclosure supports up to 12 1U half-wide compute nodes, while the iDataPlex chassis can house only two 1U half-deep compute nodes. This configuration allows NeXtScale System to provide more flexibility to the user and mix different types of nodes with different form factors in the chassis.

The NeXtScale n1200 Enclosure fits in the 42U 1100mm Enterprise V2 Dynamic Rack, but it also fits in many standard 19-inch racks. Although iDataPlex also can be installed in a standard 19-inch rack, the use of iDataPlex racks and different floor layouts was required to use its high-density capability. NeXtScale System brings more flexibility by allowing users to use standard 19-inch racks and does not require a special data center layout, which does not affect customer best practices and policies. This flexibility also allows the Lenovo NeXtScale System to achieve greater data center density when it is used among other standard 19-inch racks.

As with iDataPlex servers, NeXtScale servers support S3 mode. S3 allows systems to come back into full production from low-power state much quicker than a traditional power-on. In fact, cold start normally takes about 270 seconds; with S3, it takes only 45 seconds. When you know that a system is not to be used because of time of day or state of job flow, you can send it into a low-power state to save power and, when needed, bring it back online quickly.

Table 2-1 compares the features of NeXtScale System to those features of iDataPlex.

Table 2-1 Comparing NeXtScale System to iDataPlex

Feature	iDataPlex	NeXtScale	Comments
Form factor	Unique rack 1200 mm x 600 mm	Standard rack 600 mm x 1100 mm	NeXtScale System allows for lower-cost racks and customer's racks.
Density in a standard 42U rack	Up to 42 servers	Up to 84 servers (72 with space for switches)	NeXtScale System can provide up to twice the server density when both types of servers are installed in a standard 42U rack.
Density in two consecutive floor tiles ^a	84 servers 8 ToR switches (iDataPlex rack)	144 servers 12 ToR switches (two standard racks next to each other)	NeXtScale can provide up to 71% more servers per row when top-of-rack (ToR) switches are used.
Density/400 sq. ft. (with Rear Door Heat Exchanger)	1,680 servers 84 servers/iDataPlex rack; four rows of five racks	2,160 servers 72 servers/standard rack; three rows of 10 racks	10x10 floor tiles A 28% density increase because of standard rack layout.
Power/tile (front)	22 kW maximum 15 kW typical 42 servers + switches	37 kW maximum 25 kW typical 72 servers + switches	Similar power/server
GPU support	Two GPUs per server in 2U	Two GPUs per server in 1U effective space	GPU tray + base node = 1U. NeXtScale System has twice the density of iDataPlex.
Direct attached storage	None ^b	Other storage-rich offerings include eight drives in 1U effective space	More flexibility with NeXtScale storage plan.
Direct water cooling	Available Now	NeXtScale System design supports water cooling	Opportunity to optimize cost with NeXtScale.

a. Here we compare the density of servers that can be fitted in a single row of racks while top-of-rack switches are used. We use a single iDataPlex rack for iDataPlex servers that is 1200 mm wide, and we compare it with two standard racks for NeXtScale servers that also are 1200 mm wide.

b. None are available with Intel E5-2600 v2 series processor.

2.4 NeXtScale System versus Flex System

Although NeXtScale System and Flex System are both “blade” architectures, they are different in their approach, as shown in Figure 2-4 on page 17 and Table 2-2 on page 17. The key here is to understand the client philosophy. Features and approaches in Flex System and NeXtScale System appeal to different kinds of users.

Flex System has a wide network environment of compatible servers (x86 and POWER), switches, and storage offerings. It is aimed at clients that want an optimized and integrated solution that delivers better performance than competitive offerings.

NeXtScale System is an x86-only architecture and its aim is to be the best available server for clients that require a scale-out infrastructure. NeXtScale System uses industry-standard components, including I/O cards and top-of-rack networking switches, for flexibility of choice and ease of adoption.

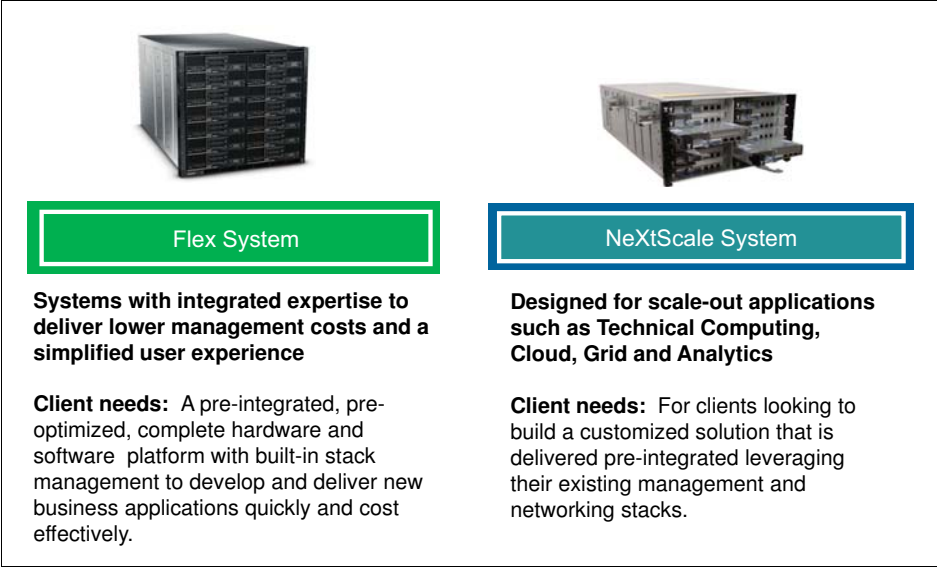


Figure 2-4 Comparing Flex System with NeXtScale System

Table 2-2 Flex System and NeXtScale System Differentiations

Flex System differentiation	NeXtScale System differentiation
Integrated design offers flexibility (Flex System) and factory-integration (PureFlex System).	Flexibility of customization and integration of hardware.
10U chassis holding 14 nodes with shared power and cooling that is designed for multiple generations of technology.	6U chassis holding 12 servers with shared power and cooling for multiple generations of technology.
Heterogeneous node support: x86 and POWER	x86 node support
Full hardware redundancy; no single point of failure.	Designed for software redundancy; clustered approach.
Integrated switching, storage, and management in the chassis. All network connections are made via midplane. Support for 1 Gb/10 Gb Ethernet, FCoE, 8 Gb/16 Gb FC, and QDR/FDR InfiniBand. Custom form factor switches and adapters.	No Integrated switching in the chassis. Cables are routed to top of rack. Multiple brands of rack-based switches are supported: 1 GB/10 GB/40 Gb Ethernet, InfiniBand QDR/FDR, and 8 GB/16 GB Fibre Channel all installed external to the chassis. Standard PCI adapters and ToR switches.
Integrated shared storage optional with the Flex System V7000 Storage Node.	Integrated direct attached storage optional with the native expansion tray.
Unified management via the Flex System Manager for all physical and virtual resources in chassis.	No unified management tools; management of node/ storage/switching handled via independent open standard, vendor-independent toolkits.

2.5 NeXtScale System versus rack-mounted servers

Although the NeXtScale System compute nodes are included in a chassis, this chassis is there to provide shared power and cooling only.

As a consequence, the approach to design, buy, or upgrade a solution that is based on NeXtScale System or regular 1U/2U rack-mounted servers is similar. In both cases, the full solution relies on separate networking components and can integrate seamlessly in an infrastructure.

The choice between NeXtScale System and rack-mountable servers is first driven by application requirements because each form factor brings various advantages and limitations for particular workloads.

The 1U/2U rack-mounted servers and the NeXtScale System feature the following main capabilities differences:

- ▶ Quantity of memory per node
- ▶ Quantity of PCIe slots per node
- ▶ Quantity of drives per node
- ▶ Support for high energy power consumption adapters, such as, GPU and co-processors

For small installations or for servers that require a large amount of memory or a large number of PCIe adapters, the rack-mounted servers is the system of choice.

For medium to large installation of nodes that require up to 256 GB, NeXtScale System is the system of choice. It allows the users to reduce their initial cost of acquisition and their operating cost through higher density (up to 4X compared to 2U servers) and higher energy efficiency (because of the shared power and cooling infrastructure).

NeXtScale n1200 Enclosure

The foundation on which NeXtScale System is built is the NeXtScale n1200 Enclosure.

Providing shared, high-efficiency power and cooling for up to 12 compute nodes, this chassis scales with your business needs. Adding compute, storage, or acceleration capability is as simple as adding nodes to the chassis. There is no built-in networking or switching capabilities, which requires no chassis-level management beyond power and cooling.

This chapter includes the following topics:

- ▶ 3.1, “Overview” on page 20
- ▶ 3.2, “Standard chassis models” on page 22
- ▶ 3.3, “Supported compute nodes” on page 22
- ▶ 3.4, “Power supplies” on page 25
- ▶ 3.5, “Fan modules” on page 28
- ▶ 3.7, “Midplane” on page 29
- ▶ 3.8, “Fan and Power Controller” on page 30
- ▶ 3.9, “Power management” on page 34
- ▶ 3.10, “Specifications” on page 37

3.1 Overview

The NeXtScale n1200 Enclosure is a 6U next-generation dense server platform with integrated Fan and Power Controller. The n1200 enclosure efficiently powers and cools up to 12 1U half-wide compute nodes, with which clients can install in a standard 42U 19-inch rack that is twice the number of servers per rack-U space that is compared to traditional 1U rack servers.



Figure 3-1 NeXtScale n1200 Enclosure with 12 compute nodes

The founding principle behind NeXtScale System is to allow clients to adopt this new hardware with minimal or no changes to their data center infrastructure, management tools, protocols, and best practices.

The enclosure looks similar to an BladeCenter or Flex System chassis, but it is different as there is no consolidated management interface or integrated switching. An exploded view of the components of the chassis is shown in Figure 3-2.

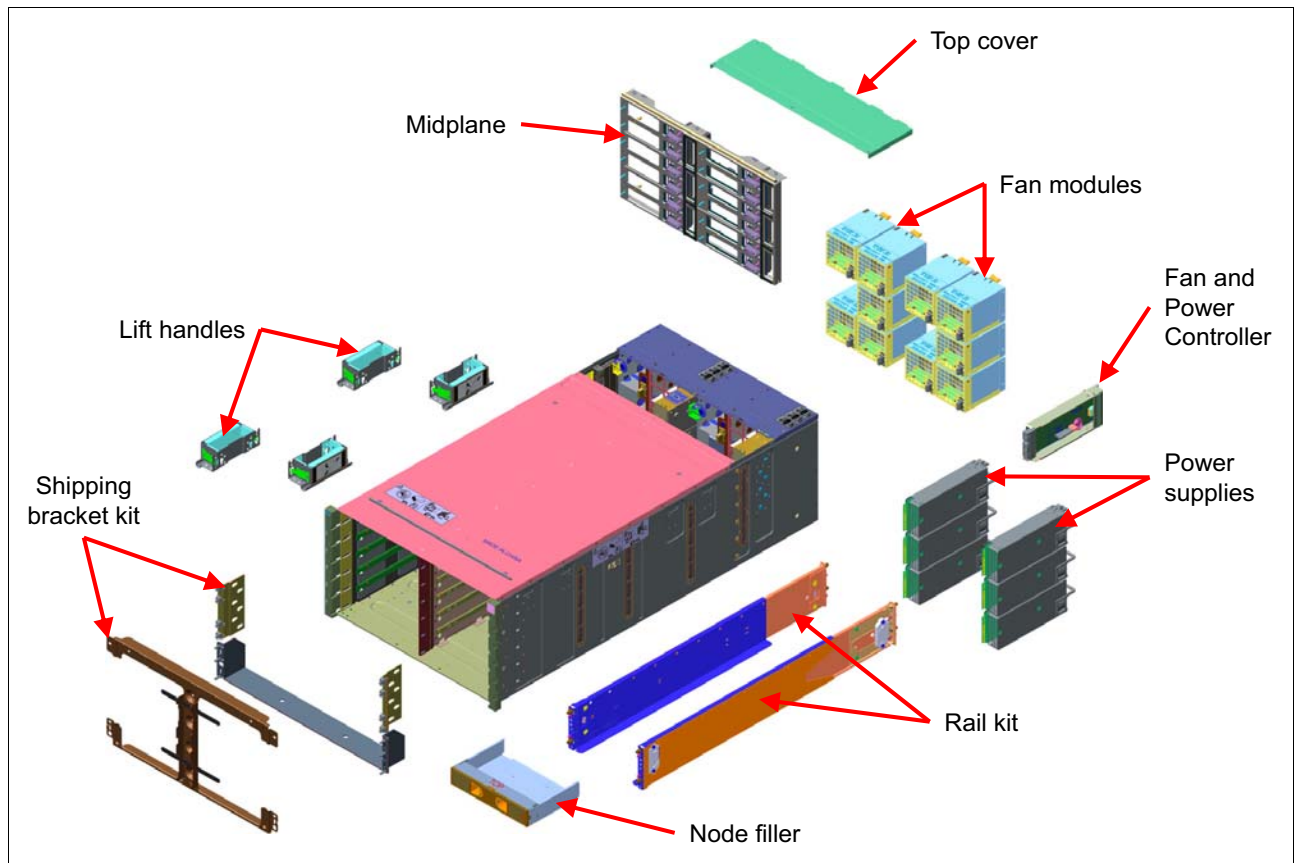


Figure 3-2 NeXtScale n1200 Enclosure components

The NeXtScale n1200 Enclosure includes the following components:

- ▶ Up to 12 compute nodes
- ▶ Six power supplies, each separately powered
- ▶ A total of 10 fan modules in two cooling zones
- ▶ One Fan and Power Controller

3.1.1 Front components

The NeXtScale n1200 Enclosure supports up to 12 1U half-wide compute nodes, as shown in Figure 3-3.

All compute nodes are front accessible with front cabling as shown in Figure 3-3. From this angle, the chassis looks to be simple because it was designed to be simple, low-cost, and efficient.

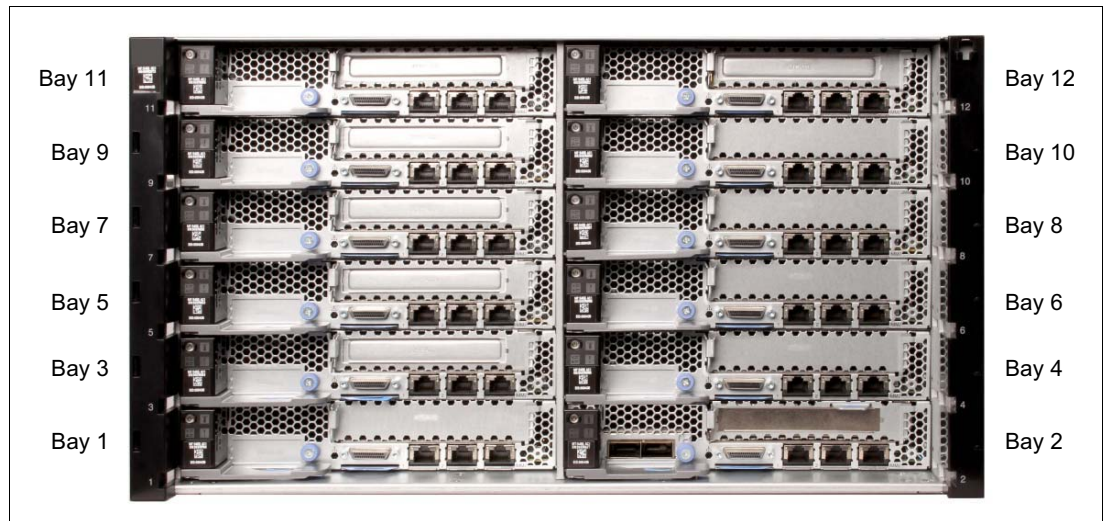


Figure 3-3 NeXtScale n1200 Enclosure front view with 12 compute nodes

This new enclosure supports dense, high-performance compute nodes, and expanded compute nodes with more I/O slots for adapters, GPUs, and coprocessors or other drive bays. As a result, clients can access some powerful IT inside a simple and cost-effective base compute node that can be expanded to create rich and dense storage or acceleration solutions without the need for any exotic components, midplanes, or high-cost connectors.

3.1.2 Rear components

The n1200 provides shared high-efficiency power supplies and fan modules. As with BladeCenter and Flex System, the NeXtScale System compute nodes connect to a midplane, but this connection is for power and control only; the midplane does not provide any I/O connectivity.

Figure 3-4 shows the major components that are accessible from the rear of the chassis.

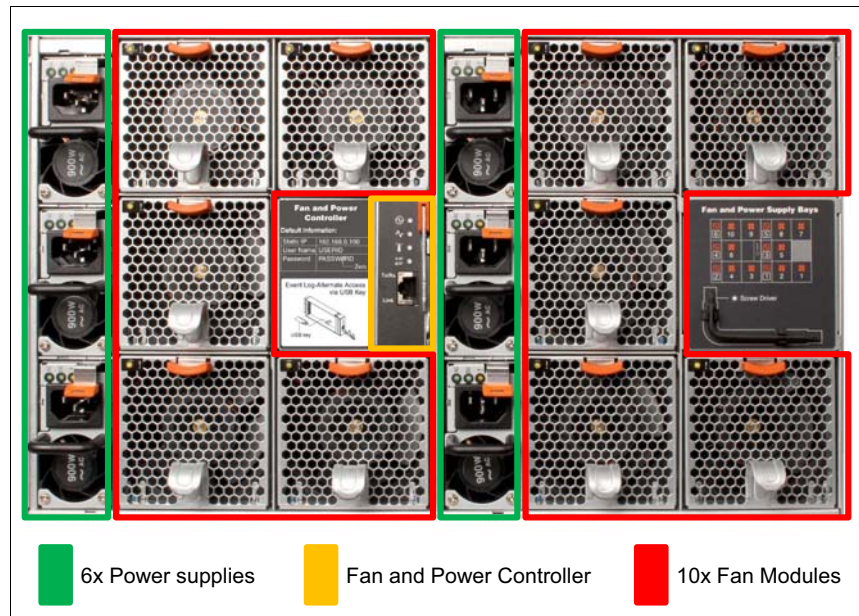


Figure 3-4 NeXtScale n1200 Enclosure rear view

At the rear of the chassis, the following types of components are accessible:

- Power supplies

The NeXtScale n1200 Enclosure has a six-power supply design, all in one power domain. This configuration allows clients with 100 V - 240 V utility power (in single or three-phase) to power up the chassis by using their available infrastructure. For three-phase power, the phases are split in the PDU for single phase input to the chassis power supplies.

For more information about the power supplies, see 3.4, “Power supplies” on page 25.

- Fan modules

Also shared in the chassis are 10 80 mm fan modules, five in each of the two cooling zones.

The fan modules and PSUs in the chassis provide shared power and cooling for all the installed nodes by using fewer components and with less power than traditional systems.

For more information about the fan modules, see 3.5, “Fan modules” on page 48.

- Fan and Power Controller

The Fan and Power Controller (FPC) module is the management device for the chassis and as its name implies, controls the power and cooling features of the enclosure.

For more information about the FPC, see 3.8, “Fan and Power Controller” on page 30.

You might notice that the enclosure does not contain space for network switches. All I/O is routed directly out of the servers to top-of-rack switches. This configuration provides choice and flexibility and keeps the NeXtScale n1200 Enclosure flexible and low-cost.

3.1.3 Fault tolerance features

The chassis implements a fault-tolerant design. The following components in the chassis enable continued operation if one of the components fails:

- Power supplies

The power supplies support a single power domain that provides DC power to all of the chassis components. If a power supply fails, the other power supplies can continue to provide power.

Power policies: The power management policy that you implemented for the chassis determines the affect on chassis operation if there is a power supply failure. Power policies can be N+N, N+1, or no redundancy. Power policies are managed by the FPC.

- Fan modules

The fan modules provide cooling to all of the chassis components. The power supplies have their own fans to provide cooling. Each fan module features a dual rotor (blade) dual motor fan. One of the motors within the fan module can fail and the remaining continues to operate. If a fan fails, the chassis continues operating and the remaining fans increase in speed to compensate.

- FPC

The FPC enables the Integrated Management Module to monitor the fans and control fan speed. If the FPC fails, the enclosure fans ramp up to maximum, but all systems continue to operate by using the power management policy.

3.2 Standard chassis models

The standard chassis models are listed in Table 3-1. The chassis is also available via the configure-to-order (CTO) process.

Table 3-1 Standard enclosure models

Model	Description	Fan Modules (standard/max)	Power Supplies (standard/max)
5456-A2x	NeXtScale n1200 Enclosure	10 x 80 mm / 10	6 x 900 W / 6
5456-A3x	NeXtScale n1200 Enclosure	10 x 80 mm / 10	2 x 1300 W / 6
5456-A4x	NeXtScale n1200 Enclosure	10 x 80 mm / 10	6 x 1300 W / 6
5456-B2x	NeXtScale n1200 Enclosure	10 x 80 mm / 10	6 x 900 W / 6
5456-B3x	NeXtScale n1200 Enclosure	10 x 80 mm / 10	2 x 1300 W / 6
5456-B4x	NeXtScale n1200 Enclosure	10 x 80 mm / 10	6 x 1300 W / 6

The NeXtScale n1200 Enclosure ships with the following items:

- ▶ Rail kit
- ▶ One Console breakout cable, also known as a KVM dongle (part number 00Y8366)
- ▶ A Torx-8 (T8) screwdriver for use with components (such as the drive cage) which is mounted on the rear of the chassis
- ▶ One AC power cord for each power supply that is installed, 1.5 m 10A, IEC320 C14 to C13 (part number 39Y7937)

Models 5456-Axx include the FPC and include IBM signed firmware. Models 5456-Bxx also include the FPC2 and have Lenovo signed firmware.

3.3 Supported compute nodes

The NeXtScale n1200 Enclosure supports the NeXtScale nx360 M4 and the NeXtScale nx360 M5. The number of compute nodes that can be powered on depends on the following factors:

- ▶ The power supply and power policy that is selected (N+N, N+1, or no redundancy)
- ▶ The AC or DC input voltage
- ▶ The components that are installed in each compute node (such as processor, memory, drives, and PCIe adapters)
- ▶ UEFI settings

To size for a specific configuration, you can use the Power Configurator that is available at this website:

<http://ibm.com/support/entry/portal/docdisplay?lnv0-pwrconf>

The following number of nodes can be operated with no performance compromise within the chassis depending on the power policy required. The tables use the following conventions:

- ▶ A green cell means that the chassis can be filled with nodes up to the maximum number that is supported in the chassis (that is, 12 nodes without GPU Trays installed, 6 servers with GPU Trays installed).
- ▶ A yellow cell means that the maximum number of nodes that the chassis can hold is fewer than the total available bays. Other bays in the chassis must remain empty.

Consider the following points regarding the tables:

- ▶ Oversubscription (OVS) of the power system allows for more efficient use of the available system power. By using oversubscription, users can make the most of the extra power from the redundant power supplies when the power supplies are in healthy condition.
- ▶ OVS and Power supply redundancy options are set via one of the available user interfaces to the Fan and Power Controller in the chassis.

Note: Some cells indicate two numbers (for example “5 + 1”), which indicates the following support for a mixture of nodes with and without the GPU Tray:

- ▶ First number: Number of nodes with a GPU Tray attached and two GPUs installed.
- ▶ Second number: Number of nodes without a GPU Tray attached.

For example, “5 + 1” means supported combination is five nodes with the GPU Tray attached, plus one node without a GPU Tray attached. In such a configuration, the one remaining server bay in the chassis must remain empty.

3.3.1 nx360 M4 node support

This section describes the maximum number of nodes that can be supported under various configurations.

The nx360 M4 features the following node support tables:

- ▶ 1300 W power supply:
 - 200 - 240 V AC input, no GPU Trays; see Table 3-2
 - 200 - 240 V AC input, GPU Trays with 130 W GPUs; see Table 3-3 on page 27
 - 200 - 240 V AC input, GPU Trays with 225 W GPUs; see Table 3-4 on page 28
 - 200 - 240 V AC input, GPU Trays with 235 W GPUs; see Table 3-5 on page 29
 - 200 - 240 V AC input, GPU Trays with 300 W GPUs; see Table 3-6 on page 29
- ▶ 900 W power supply:
 - 200 - 240 V AC input, no GPU Trays; see Table 3-7 on page 30
 - 100 - 127 V AC input, no GPU Trays; see Table 3-8 on page 31

Table 3-2 shows the supported quantity of compute nodes with six 1300 W power supplies installed in the chassis.

Table 3-2 Number of supported compute nodes (200 - 240 V AC Input, with 6 x 1300 W PSUs)

Compute nodes		Power policy: 6 x 1300 W power supplies			
CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS ^a	N+1	N+N	N+N with OVS ^a
50 W	1	12	12	12	12
	2	12	12	12	12
60 W	1	12	12	12	12
	2	12	12	12	12
70 W	1	12	12	12	12
	2	12	12	12	12
80 W	1	12	12	12	12
	2	12	12	12	12
95 W	1	12	12	12	12
	2	12	12	10	12
115 W	1	12	12	12	12
	2	12	12	8	12

Compute nodes		Power policy: 6 x 1300 W power supplies			
CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS ^a	N+1	N+N	N+N with OVS ^a
130 W	1	12	12	12	12
	2	12	12	7	11

a. Oversubscription (OVS) of the power system allows for more efficient use of the available system power.

Table 3-3 shows the supported quantity of compute nodes with GPU trays attached and two 130 W GPUs installed. Only 1300 W power supplies are supported and only 200 - 240 V AC input is supported.

Table 3-3 Number of supported compute nodes with two 130 W GPUs (200 - 240 V Input, with 6 x 1300 W PSUs)

Compute nodes		Power policy: 6 x 1300 W power supplies			
CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS ^a	N+1	N+N	N+N with OVS ^a
50 W	1	6	6	6	6
	2	6	6	6	6
60 W	1	6	6	6	6
	2	6	6	6	6
70 W	1	6	6	6	6
	2	6	6	6	6
80 W	1	6	6	6	6
	2	6	6	6	6
95 W	1	6	6	6	6
	2	6	6	5 + 1 ^b	6
115 W	1	6	6	6	6
	2	6	6	5	6
130 W	1	6	6	5 + 1 ^b	6
	2	6	6	4 + 1 ^b	5 + 1 ^b

a. Oversubscription (OVS) of the power system allows for more efficient use of the available system power.

b. See shaded box on page 26.

Table 3-4 shows the quantity of supported compute nodes with GPU trays attached and two 225 W GPUs installed. Only 1300 W power supplies are supported and only 200 - 240 V AC input is supported.

Table 3-4 Number of supported compute nodes with two 225 W GPUs (200 - 240 V Input, with 6 x 1300 W PSUs)

Compute nodes		Power policy: 6 x 1300 W power supplies			
CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS ^a	N+1	N+N	N+N with OVS ^a
50 W	1	6	6	5 + 1 ^b	6
	2	6	6	5	6
60 W	1	6	6	5	6
	2	6	6	4 + 1 ^b	5 + 1 ^b
70 W	1	6	6	5	6
	2	6	6	4 + 1 ^b	5 + 1 ^b
80 W	1	6	6	5	6
	2	6	6	4 + 1 ^b	5 + 1 ^b
95 W	1	6	6	4 + 2 ^b	6
	2	6	6	4	5
115 W	1	6	6	4 + 1 ^b	5 + 1 ^b
	2	6	6	3 + 1 ^b	4 + 1 ^b
130 W	1	6	6	4 + 1 ^b	5
	2	6	6	3 + 1 ^b	4 + 1 ^b

a. Oversubscription (OVS) of the power system allows for more efficient use of the available system power.

b. See shaded box on page 26.

Table 3-5 shows the supported quantity of compute nodes with GPU trays attached and two 235 W GPUs installed. Only 1300 W power supplies are supported and only 200 - 240 V AC input is supported.

Table 3-5 Number of supported compute nodes with two 235 W GPUs (200 - 240 V Input, with 6 x 1300 W PSUs)

Compute nodes		Power policy: 6 x 1300 W power supplies			
CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS ^a	N+1	N+N	N+N with OVS ^a
50 W	1	6	6	5 + 1 ^b	6
	2	6	6	4 + 1 ^b	6
60 W	1	6	6	5	6
	2	6	6	4 + 1 ^b	5 + 1 ^b
70 W	1	6	6	5	6
	2	6	6	4 + 1 ^b	5 + 1 ^b
80 W	1	6	6	5	6
	2	6	6	4 + 1 ^b	5 + 1 ^b
95 W	1	6	6	4 + 2 ^b	5 + 1 ^b
	2	6	6	4	5
115 W	1	6	6	4 + 1 ^b	5 + 1 ^b
	2	6	6	3 + 1 ^b	4 + 1 ^b
130 W	1	6	6	4	5
	2	6	6	3 + 1 ^b	4

a. Oversubscription (OVS) of the power system allows for more efficient use of the available system power.

b. See shaded box on page 26.

Table 3-6 shows the supported quantity of compute nodes with GPU trays attached and two 300 W GPUs installed. Only 1300 W power supplies are supported and only 200 - 240 V AC input is supported. The supported 300 W GPU is Intel Xeon Phi 7120P, part number 00J6162.

Table 3-6 Number of supported compute nodes with two 300 W GPUs (200 - 240 V Input, with 6 x 1300 W PSUs)

Compute nodes		Power policy: 6 x 1300 W power supplies			
CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS ^a	N+1	N+N	N+N with OVS ^a
50 W	1	6	6	4 + 2 ^b	5 + 1 ^b
	2	6	6	4	5
60 W	1	6	6	4	5
	2	6	6	3 + 2 ^b	4 + 2 ^b
70 W	1	6	6	4	5
	2	6	6	3 + 2 ^b	4 + 2 ^b

Compute nodes		Power policy: 6 x 1300 W power supplies			
CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS ^a	N+1	N+N	N+N with OVS ^a
80 W	1	6	6	4	5
	2	6	6	3 + 2 ^b	4 + 2 ^b
95 W	1	6	6	4	4 + 2 ^b
	2	6	6	3 + 1 ^b	4 + 1 ^b
115 W	1	6	6	3 + 2 ^b	4 + 2 ^b
	2	6	5 + 1 ^b	3	3 + 2 ^b
130 W	1	6	6	3 + 2 ^b	4 + 1 ^b
	2	6	5 + 1 ^b	3	3 + 2 ^b

a. OVS (oversubscription) of the power system allows for more efficient use of the available system power.

b. See shaded box on page 26

Table 3-7 shows the supported quantity of compute nodes with six 900 W power supplies installed in the chassis with those power supplies connected to a 200 - 240 V (high-line) AC input.

Table 3-7 Number of supported compute nodes (200 - 240 V AC Input, with 6 x 900 W PSUs)

Compute nodes		Power policy: 6 x 900 W power supplies			
CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS ^a	N+1	N+N	N+N with OVS ^a
50 W	1	12	12	12	12
	2	12	12	11	12
60 W	1	12	12	12	12
	2	12	12	10	12
70 W	1	12	12	12	12
	2	12	12	8	11
80 W	1	12	12	11	12
	2	12	12	8	9
95 W	1	12	12	10	12
	2	12	12	6	10
115 W	1	12	12	8	10
	2	12	10	5	8
130 W	1	12	12	7	9
	2	10	8	4	7

a. Oversubscription (OVS) of the power system allows for more efficient use of the available system power.

Table 3-8 shows the supported quantity of compute nodes with six 900 W power supplies installed in the chassis with those power supplies connected to a 100 - 127 V (low-line) AC input.

Table 3-8 Number of supported compute nodes (100 - 1127 V AC Input, with 6 x 900 W PSUs)

Compute nodes		Power policy: 6 x 900 W power supplies			
CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS ^a	N+1	N+N	N+N with OVS ^a
50 W	1	12	12	9	11
	2	12	12	6	10
60 W	1	12	12	7	9
	2	12	9	5	7
70 W	1	12	12	7	9
	2	12	9	5	7
80 W	1	12	12	6	8
	2	10	9	5	7
95 W	1	12	11	6	7
	2	9	7	4	6
115 W	1	11	9	5	6
	2	7	6	3	5
130 W	1	9	8	4	5
	2	6	5	3	4

a. Oversubscription (OVS) of the power system allows for more efficient use of the available system power.

3.3.2 nx360 M5 node support

This section describes the maximum number of nodes that can be supported under various configurations.

The tables in this section are as follows:

- ▶ 1500W power supplies, 200-240V AC input, and no GPUs
 - Table 3-9 on page 32 - 1500 W power supplies, 200-240V AC input, no GPUs
- ▶ 1500W power supplies, 200-240V AC input, with 1U PCIe Native Expansion Tray
 - Table 3-10 on page 33 - 1500 W power supplies, with two 130 W GPUs
 - Table 3-11 on page 33 - 1500 W power supplies, with two 225 W GPUs
 - Table 3-12 on page 34 - 1500 W power supplies, with two 235 W GPUs
 - Table 3-13 on page 34 - 1500 W power supplies, with two 300 W GPUs
- ▶ 1500W power supplies, 200-240V AC input, with 2U PCIe Native Expansion Tray
 - Table 3-14 on page 35 - 1500 W power supplies, with four 130 W GPUs
 - Table 3-15 on page 35 - 1500 W power supplies, with four 225 W GPUs
 - Table 3-16 on page 36 - 1500 W power supplies, with four 235 W GPUs
 - Table 3-17 on page 36 - 1500 W power supplies, with four 300 W GPUs

- ▶ 1300W power supplies 200-240V AC input, and no GPUs
 - Table 3-18 on page 37 - 1300 W power supplies, no GPUs
- ▶ 1300W power supplies (200-240V AC input) with 1U PCIe Native Expansion Tray
 - Table 3-19 on page 38 - 1300 W power supplies, with two 130 W GPUs
 - Table 3-20 on page 38 - 1300 W power supplies, with two 225 W GPUs
 - Table 3-21 on page 39 - 1300 W power supplies, with two 235 W GPUs
 - Table 3-22 on page 39 - 1300 W power supplies, with two 300 W GPUs
- ▶ 1300W power supplies (200-240V AC input) with 2U PCIe Native Expansion Tray
 - Table 3-23 on page 40 - 1300 W power supplies, with four 130 W GPUs
 - Table 3-24 on page 40 - 1300 W power supplies, with four 225 W GPUs
 - Table 3-25 on page 41 - 1300 W power supplies, with four 235 W GPUs
 - Table 3-26 on page 41 - 1300 W power supplies, with four 300 W GPUs
- ▶ 900W power supplies
 - Table 3-27 on page 42 - 900 W power supplies, 200-240V AC input, no GPUs
 - Table 3-28 on page 43 - 900 W power supplies, 100-127V AC input, no GPUs
 - Table 3-29 on page 44 - 900 W power supplies, -48 V DC input, no GPUs

Chassis with six 1500 W power supplies

Table 3-9 shows the supported quantity of compute nodes with six 1500 W power supplies installed in the chassis.

Table 3-9 Number of compute nodes that are supported (200 - 240 V AC Input, with 6 x 1500 W PSUs)

CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS	N+1	N+N	N+N with OVS
55 W	1	12	12	12	12
	2	12	12	12	12
65 W	1	12	12	12	12
	2	12	12	11	12
85 W	1	12	12	12	12
	2	12	12	9	10
90 W	1	12	12	12	12
	2	12	12	8	10
105 W	1	12	12	11	12
	2	12	12	7	9
120 W	1	12	12	10	12
	2	12	11	6	8
135 W	1	12	12	9	11
	2	12	10	6	7
145 W	1	12	12	8	10
	2	12	10	5	7

Table 3-10 shows the supported quantity of compute nodes with 1U GPU trays attached and two 130 W GPUs installed. Compute nodes are installed in a chassis with 1500 W power supplies with 200 - 240 V AC input.

Table 3-10 Number of compute nodes that are supported each with two 130 W GPUs installed in the 1U PCIe Native Expansion Tray (200 - 240 V AC Input, with 6 x 1500 W PSUs)

CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS	N+1	N+N	N+N with OVS
55 W	2	6	6	6	6
65 W	2	6	6	5 + 1	6
85 W	2	6	6	5	6
90 W	2	6	6	5	6
105 W	2	6	6	5	6
120 W	2	6	6	4 + 1	5 + 1
135 W	2	6	6	4	5
145 W	2	6	6	4	5

Table 3-11 shows the supported quantity of compute nodes with 1U GPU trays attached and two 225 W GPUs installed. Compute nodes are installed in a chassis with 1500 W power supplies with 200 - 240 V AC input.

Table 3-11 Number of compute nodes that are supported each with two 225 W GPUs installed in the 1U PCIe Native Expansion Tray (200 - 240 V AC Input, with 6 x 1500 W PSUs)

CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS	N+1	N+N	N+N with OVS
55 W	2	6	6	4 + 1	5 + 1
65 W	2	6	6	4	5
85 W	2	6	6	4	5
90 W	2	6	6	4	5
105 W	2	6	6	3 + 1	4 + 1
120 W	2	6	6	3 + 1	4 + 1
135 W	2	6	6	3 + 1	4
145 W	2	6	6	3 + 1	4

Table 3-12 shows the supported quantity of compute nodes with 1U GPU trays attached and two 235 W GPUs installed. Compute nodes are installed in a chassis with 1500 W power supplies with 200 - 240 V AC input.

Table 3-12 Number of compute nodes that are supported each with two 235 W GPUs installed in the 1U PCIe Native Expansion Tray (200 - 240 V AC Input, with 6 x 1500 W PSUs) CPU TDP

Number of CPUs	Non-redundant or N+1 with OVS	N+1	N+N	N+N with OVS	
55 W	2	6	6	4 + 1	5
65 W	2	6	6	4	5
85 W	2	6	6	4	5
90 W	2	6	6	4	4 + 1
105 W	2	6	6	3 + 1	4 + 1
120 W	2	6	6	3 + 1	4 + 1
135 W	2	6	6	3 + 1	4
145 W	2	6	6	3	4

Table 3-13 shows the supported quantity of compute nodes with 1U GPU trays attached and two 300 W GPUs installed. Compute nodes are installed in a chassis with 1500 W power supplies with 200 - 240 V AC input.

Table 3-13 Number of compute nodes that are supported each with two 300 W GPUs installed in the 1U PCIe Native Expansion Tray (200 - 240 V AC Input, with 6 x 1500 W PSUs)

CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS	N+1	N+N	N+N with OVS
55 W	2	6	6	3 + 2	4 + 1
65 W	2	6	6	3 + 1	4 + 1
85 W	2	6	6	3 + 1	4
90 W	2	6	6	3 + 1	4
105 W	2	6	5 + 1	3	4
120 W	2	6	5 + 1	3	4
135 W	2	6	5	3	3 + 1
145 W	2	6	5	3	3 + 1

Table 3-14 shows the supported quantity of compute nodes with 2U GPU trays attached and four 130 W GPUs installed. Compute nodes are installed in a chassis with 1500 W power supplies with 200 - 240 V AC input.

Table 3-14 Number of compute nodes that are supported each with four 130 W GPUs installed in the 2U PCIe Native Expansion Tray (200 - 240 V AC Input, with 6 x 1500 W PSUs)

CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS	N+1	N+N	N+N with OVS
55 W	2	4	4	4	4
65 W	2	4	4	4	4
85 W	2	4	4	4	4
90 W	2	4	4	4	4
105 W	2	4	4	4	4
120 W	2	4	4	4	4
135 W	2	4	4	4	4
145 W	2	4	4	3 + 1	4

Table 3-15 shows the supported quantity of compute nodes with 2U GPU trays attached and four 225 W GPUs installed. Compute nodes are installed in a chassis with 1500 W power supplies with 200 - 240 V AC input.

Table 3-15 Number of compute nodes that are supported each with four 225 W GPUs installed in the 2U PCIe Native Expansion Tray (200 - 240 V AC Input, with 6 x 1500 W PSUs)

CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS	N+1	N+N	N+N with OVS
55 W	2	4	4	3 + 1	4
65 W	2	4	4	3	3 + 3
85 W	2	4	4	3	3 + 2
90 W	2	4	4	3	3 + 2
105 W	2	4	4	3	3 + 2
120 W	2	4	4	2 + 2	3 + 1
135 W	2	4	4	2 + 2	3 + 1
145 W	2	4	4	2 + 2	3 + 1

Table 3-16 shows the supported quantity of compute nodes with 2U GPU trays attached and four 235 W GPUs installed. Compute nodes are installed in a chassis with 1500 W power supplies with 200 - 240 V AC input.

Table 3-16 Number of compute nodes that are supported each with four 235 W GPUs installed in the 2U PCIe Native Expansion Tray (200 - 240 V AC Input, with 6 x 1500 W PSUs) CPU

TDP	Number of CPUs	Non-redundant or N+1 with OVS	N+1	N+N	N+N with OVS
55 W	2	4	4	3	3 + 3
65 W	2	4	4	3	3 + 2
85 W	2	4	4	3	3 + 2
90 W	2	4	4	3	3 + 2
105 W	2	4	4	2 + 2	3 + 1
120 W	2	4	4	2 + 2	3 + 1
135 W	2	4	4	2 + 2	3 + 1
145 W	2	4	4	2 + 1	3

Table 3-17 shows the supported quantity of compute nodes with 2U GPU trays attached and four 300 W GPUs installed. Compute nodes are installed in a chassis with 1500 W power supplies with 200 - 240 V AC input.

Table 3-17 Number of compute nodes that are supported each with four 300 W GPUs installed in the 2U PCIe Native Expansion Tray (200 - 240 V AC Input, with 6 x 1500 W PSUs)

CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS	N+1	N+N	N+N with OVS
55 W	2	4	4	2 + 2	3 + 1
65 W	2	4	4	2 + 2	3
85 W	2	4	4	2 + 2	3
90 W	2	4	4	2 + 1	3
105 W	2	4	4	2 + 1	3
120 W	2	4	4	2 + 1	2 + 3
135 W	2	4	4	2 + 1	2 + 2
145 W	2	4	4	2 + 1	2 + 2

Chassis with six 1300 W power supplies

Table 3-18 shows the supported quantity of compute nodes with six 1300 W power supplies installed in the chassis.

Table 3-18 Number of supported compute nodes (200 - 240 V AC Input, with 6 x 1300 W PSUs)

Compute nodes		Power policy: 6 x 1300 W power supplies			
CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS ^a	N+1	N+N	N+N with OVS ^a
55 W	1	12	12	12	12
	2	12	12	10	12
65 W	1	12	12	12	12
	2	12	12	9	11
85 W	1	12	12	12	12
	2	12	12	8	10
90 W	1	12	12	12	12
	2	12	12	7	9
105 W	1	12	12	12	12
	2	12	12	7	8
120 W	1	12	12	11	12
	2	12	11	6	8
135 W	1	12	12	11	12
	2	12	10	6	7
145 W	1	12	12	10	12
	2	12	10	5	7

a. Oversubscription (OVS) of the power system allows for more efficient use of the available system power.

Table 3-19 shows the supported quantity of compute nodes with 1U GPU trays attached and two 130 W GPUs installed. Compute nodes are installed in a chassis with 1300 W power supplies with 200 - 240 V AC input.

Table 3-19 Number of supported compute nodes with two 130 W GPUs (200 - 240 V AC Input, with 6 x 1300 W PSUs)

Compute nodes		Power policy: 6 x 1300 W power supplies			
CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS ^a	N+1	N+N	N+N with OVS ^a
55 W	2	6	6	6	6
65 W	2	6	6	5 + 1 ^b	6
85 W	2	6	6	5	6
90 W	2	6	6	5	6
105 W	2	6	6	5	6
120 W	2	6	6	4 + 1 ^b	5 + 1 ^b
135 W	2	6	6	4	5
145 W	2	6	6	4	5

a. Oversubscription (OVS) of the power system allows for more efficient use of the available system power.

b. See shaded box on page 26.

Table 3-20 shows the supported quantity of compute nodes with 1U GPU trays attached and two 225 W GPUs installed. Compute nodes are installed in a chassis with 1300 W power supplies with 200 - 240 V AC input.

Table 3-20 Number of supported compute nodes with two 225 W GPUs (200 - 240 V AC Input, with 6 x 1300 W PSUs)

Compute nodes		Power policy - 6 x 1300 W power supplies			
CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS ^a	N+1	N+N	N+N with OVS ^a
55 W	2	6	6	4 + 1 ^b	5 + 1 ^b
65 W	2	6	6	4	5
85 W	2	6	6	4	5
90 W	2	6	6	4	5
105 W	2	6	6	3 + 1 ^b	4 + 1 ^b
120 W	2	6	6	3 + 1 ^b	4 + 1 ^b
135 W	2	6	6	3 + 1 ^b	4
145 W	2	6	6	3 + 1 ^b	4

a. Oversubscription (OVS) of the power system allows for more efficient use of the available system power.

b. See shaded box on page 26.

Table 3-21 shows the supported quantity of compute nodes with 1U GPU trays attached and two 235 W GPUs installed. Compute nodes are installed in a chassis with 1300 W power supplies with 200 - 240 V AC input.

Table 3-21 Number of supported compute nodes with two 235 W GPUs (200 - 240 V AC Input, with 6 x 1300 W PSUs)

Compute nodes		Power policy: 6 x 1300 W power supplies			
CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS ^a	N+1	N+N	N+N with OVS ^a
55 W	2	6	6	4 + 1 ^b	5
65 W	2	6	6	4	5
85 W	2	6	6	4	5
90 W	2	6	6	4	4 + 1 ^b
105 W	2	6	6	3 + 1 ^b	4 + 1 ^b
120 W	2	6	6	3 + 1 ^b	4 + 1 ^b
135 W	2	6	6	3 + 1 ^b	4
145 W	2	6	6	3	4

a. Oversubscription (OVS) of the power system allows for more efficient use of the available system power.

b. See shaded box on page 26.

Table 3-22 shows the supported quantity of compute nodes with 1U GPU trays attached and two 300 W GPUs installed. Compute nodes are installed in a chassis with 1300 W power supplies with 200 - 240 V AC input.

Table 3-22 Number of supported compute nodes with two 300 W GPUs (200 - 240 V AC Input, with 6 x 1300 W PSUs)

Compute nodes		Power policy: 6 x 1300 W power supplies			
CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS ^a	N+1	N+N	N+N with OVS ^a
55 W	2	6	6	3 + 2 ^b	4 + 1 ^b
65 W	2	6	6	3 + 1 ^b	4 + 1 ^b
85 W	2	6	6	3 + 1 ^b	4
90 W	2	6	6	3 + 1 ^b	4
105 W	2	6	5 + 1 ^b	3	4
120 W	2	6	5 + 1 ^b	3	4
135 W	2	6	5	3	3 + 1 ^b
145 W	2	6	5	3	3 + 1 ^b

a. Oversubscription (OVS) of the power system allows for more efficient use of the available system power.

b. See shaded box on page 26.

Table 3-23 shows the supported quantity of compute nodes with 2U GPU trays attached and four 130 W GPUs installed. Compute nodes are installed in a chassis with 1300 W power supplies with 200 - 240 V AC input.

Table 3-23 Number of compute nodes that are supported each with four 130 W GPUs installed in the 2U PCIe Native Expansion Tray (200 - 240 V AC Input, with 6 x 1300 W PSUs)

CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS	N+1	N+N	N+N with OVS
55 W	2	4	4	4	4
65 W	2	4	4	4	4
85 W	2	4	4	3 + 1	4
90 W	2	4	4	3 + 1	4
105 W	2	4	4	3 + 1	4
120 W	2	4	4	3	4
135 W	2	4	4	3	4
145 W	2	4	4	3	4

Table 3-24 shows the supported quantity of compute nodes with 2U GPU trays attached and four 225 W GPUs installed. Compute nodes are installed in a chassis with 1300 W power supplies with 200 - 240 V AC input.

Table 3-24 Number of compute nodes that are supported each with four 225 W GPUs installed in the 2U PCIe Native Expansion Tray (200 - 240 V AC Input, with 6 x 1300 W PSUs)

CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS	N+1	N+N	N+N with OVS
55 W	2	4	4	2 + 3	3 + 1
65 W	2	4	4	2 + 2	3 + 1
85 W	2	4	4	2 + 2	3
90 W	2	4	4	2 + 2	3
105 W	2	4	4	2 + 1	3
120 W	2	4	4	2 + 1	3
135 W	2	4	4	2 + 1	3
145 W	2	4	4	2 + 1	2 + 2

Table 3-25 shows the supported quantity of compute nodes with 2U GPU trays attached and four 235 W GPUs installed. Compute nodes are installed in a chassis with 1300 W power supplies with 200 - 240 V AC input.

Table 3-25 Number of compute nodes that are supported each with four 235 W GPUs installed in the 2U PCIe Native Expansion Tray (200 - 240 V AC Input, with 6 x 1300 W PSUs) CPU

TDP	Number of CPUs	Non-redundant or N+1 with OVS	N+1	N+N	N+N with OVS
55 W	2	4	4	2 + 2	3 + 1
65 W	2	4	4	2 + 2	3 + 1
85 W	2	4	4	2 + 1	3
90 W	2	4	4	2 + 1	3
105 W	2	4	4	2 + 1	3
120 W	2	4	4	2 + 1	3
135 W	2	4	4	2 + 1	2 + 2
145 W	2	4	4	2	2 + 2

Table 3-26 shows the supported quantity of compute nodes with 2U GPU trays attached and four 300 W GPUs installed. Compute nodes are installed in a chassis with 1300 W power supplies with 200 - 240 V AC input.

Table 3-26 Number of compute nodes that are supported each with four 300 W GPUs installed in the 2U PCIe Native Expansion Tray (200 - 240 V AC Input, with 6 x 1300 W PSUs)

CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS	N+1	N+N	N+N with OVS
55 W	2	4	3 + 3	2 + 1	2 + 3
65 W	2	4	3 + 3	2	2 + 3
85 W	2	4	3 + 2	2	2 + 2
90 W	2	4	3 + 2	2	2 + 2
105 W	2	4	3 + 2	2	2 + 2
120 W	2	4	3 + 1	2	2 + 1
135 W	2	4	3 + 1	2	2 + 1
145 W	2	4	3 + 1	2	2 + 1

Chassis with six 900 W power supplies

Table 3-27 shows the supported quantity of compute nodes with six 900 W power supplies installed in the chassis with those power supplies connected to a 200 - 240 V (high-line) AC input.

Table 3-27 Number of supported compute nodes (200 - 240 V AC Input, with 6 x 900 W PSUs)

Compute nodes		Power policy: 6 x 900 W power supplies			
CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS ^a	N+1	N+N	N+N with OVS ^a
55 W	1	12	12	11	12
	2	12	11	6	8
65 W	1	12	12	10	12
	2	12	10	6	7
85 W	1	12	12	9	11
	2	11	9	5	6
90 W	1	12	12	9	11
	2	11	9	5	6
105 W	1	12	12	8	10
	2	10	8	4	5
120 W	1	12	12	7	9
	2	9	7	4	5
135 W	1	12	12	7	9
	2	8	7	4	5
145 W	1	12	12	7	8
	2	8	6	3	4

a. Oversubscription (OVS) of the power system allows for more efficient use of the available system power.

Table 3-28 shows the supported quantity of compute nodes with six 900 W power supplies installed in the chassis with those power supplies connected to a 100-127 V (low-line) AC input.

Note: When the 900 W power supply is operated at 100 - 127 V input voltage, the capacity of the power supply is reduced to 600 W.

Table 3-28 Number of supported compute nodes (100 - 127 V AC Input, with 6 x 900 W PSUs)

Compute nodes		Power policy: 6 x 900 W power supplies			
CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS ^a	N+1	N+N	N+N with OVS ^a
55 W	1	12	12	6	8
	2	9	7	4	5
65 W	1	12	11	6	8
	2	8	6	3	4
85 W	1	12	10	5	7
	2	7	6	3	4
90 W	1	12	10	5	7
	2	7	5	3	4
105 W	1	11	9	5	6
	2	6	5	2	3
120 W	1	10	8	4	6
	2	6	4	2	3
135 W	1	10	8	4	5
	2	5	4	2	3
145 W	1	9	7	4	5
	2	5	4	2	3

a. Oversubscription (OVS) of the power system allows for more efficient use of the available system power.

Table 3-29 shows the supported quantity of compute nodes with six 900 W power supplies installed in the chassis with those power supplies connected to a -48 V DC input.

Table 3-29 Number of supported compute nodes (-48 V DC Input, with 6 x 900 W PSUs)

Compute nodes		Power policy: 6 x 900 W power supplies			
CPU TDP	Number of CPUs	Non-redundant or N+1 with OVS ^a	N+1	N+N	N+N with OVS ^a
55 W	1	12	12	11	12
	2	12	11	6	8
65 W	1	12	12	10	12
	2	12	10	6	7
85 W	1	12	12	9	11
	2	11	9	5	6
90 W	1	12	12	9	11
	2	11	9	5	6
105 W	1	12	12	8	10
	2	10	8	4	5
120 W	1	12	12	7	9
	2	9	7	4	5
135 W	1	12	12	7	9
	2	8	7	4	5
145 W	1	12	12	7	8
	2	8	6	3	4

a. Oversubscription (OVS) of the power system allows for more efficient use of the available system power.

3.4 Power supplies

The NeXtScale n1200 Enclosure supports up to six high-efficiency autoranging power supplies. The standard model includes all six power supplies. Available AC power supplies are shown in Figure 3-5.



Figure 3-5 Available power supplies for NeXtScale n1200 Enclosure

Table 3-4 lists the ordering information for the supported power supplies.

Table 3-30 Power supplies

Part number	Feature code	Description	Min / Max supported	Chassis model where used
00Y8569	A41T	CFF 900 W Power Supply (80 PLUS Platinum)	6 / 6	A2x, B2x
00Y8652	A4MM	CFF 1300 W Power Supply (80 PLUS Platinum)	2 / 6	A3x, A4x, B3x, B4x
00MU774	ASYH	NeXtScale n1200 1300W Titanium Power Supply	2 / 6	-
00MU775	ASYJ	NeXtScale n1200 1500W Platinum Power Supply	2 / 6	-
00KG685	ASGJ	CFF -48 V DC 900 W Power Supply ^a	6 / 6	-

a. When the CFF -48V DC 900 W Power Supply is selected, the FPC2 (Lenovo Signed) must be selected.

The power supply options include the following features:

- ▶ Supports N+N or N+1 Power Redundancy, or Non-redundant power configurations to support higher density
- ▶ Power management controller and configured through the Fan and Power Controller
- ▶ Integrated 2500 RPM fan
- ▶ 80 PLUS Platinum or Titanium certified
- ▶ Built-in overload and surge protection

The 900 W AC power supply features the following specifications:

- ▶ Supports dual-range voltage: 100 - 240 V
- ▶ 100 - 127 (nominal) V AC; 50 or 60 Hz; 6.8 A (maximum)
- ▶ 200 - 240 (nominal) V AC; 50 or 60 Hz; 5.0 A (maximum)

The 900 W -48 V DC power supply features the following specifications:

- ▶ Supports DC voltage only: -40.5 V to -57 V
- ▶ -40.5 to -57 (nominal) V DC; 25A (maximum)

The 1300 W AC power supply features the following specifications:

- ▶ Supports high-range voltage only: 200 - 240 V
- ▶ 200 - 240 (nominal) V AC; 50 or 60 Hz; 6.9 A (maximum)

1500 W AC power supply specifications:

- ▶ Supports high-range voltage only: 200 - 240 V
- ▶ 200 - 240 (nominal) V AC; 50 or 60 Hz; 8.2 A (maximum)

200-240V only: The 1500 W AC and 1300 W AC power supplies do not support low-range voltages (100 - 127 V).

The location and numbering of the power supplies are shown in Figure 3-6.

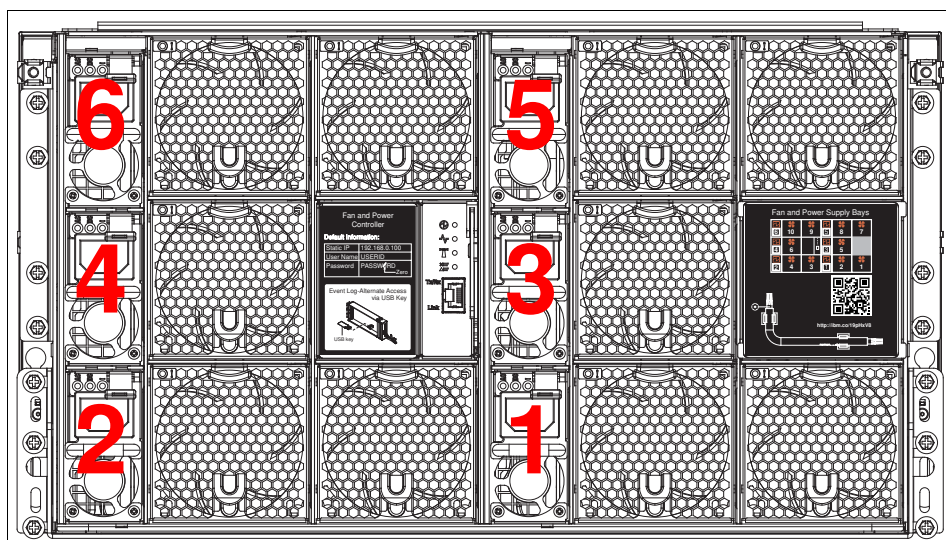


Figure 3-6 NeXtScale n1200 Enclosure rear view with power supply numbering

The power supplies that are used in NeXtScale System are hot-swap, high-efficiency 80 PLUS Platinum power supplies that are operating at 94% peak efficiency. The efficiency varies by load, as shown in Table 3-5. The 80 PLUS report is available at the following websites:

- ▶ 900 W AC Power Supply 80 PLUS report:
http://www.pluginloadsolutions.com/psu_reports/IBM_7001700-XXXX_900W_S0-571_Report.pdf
- ▶ 1300 W Power Supply 80 PLUS report:
http://www.pluginloadsolutions.com/psu_reports/IBM_700-013496-XXXX_1300W_S0-628_Report.pdf

Table 3-31 Power efficiencies at different load levels

	20% load	50% load	100% load
80 PLUS Platinum standard	90.00%	94.00%	91.00%
NeXtScale n1200 900 W AC power supply	92.00%	94.00%	91.00%
NeXtScale n1200 1300 W power supply	92.00%	94.00%	91.00%
NeXtScale n1200 900 W -48 V DC power supply ^a	92.00%	94.00%	91.00%

- a. 80 PLUS certifications do not apply to the -48 V DC power supply. The efficiencies are shown just for reference.

The 80 PLUS performance specification is for power supplies that are used within servers and computers. To meet the 80 PLUS standard, the power supply must have an efficiency of 80% or greater, at 20%, 50%, and 100% of rated load with PF of 0.9 or greater. The standard includes several grades, such as, Bronze, Silver, Gold, Platinum, and Titanium. For more information about the 80 PLUS standard, see this website:

<http://www.80PLUS.org>

The power supplies receive electrical power from a 100 V - 127 V AC or 200 V- 240 V AC power source and convert the AC input into DC outputs. The power supplies can autorange within the input voltage range.

Use with 110 V - 127 V AC: When low input voltage (100 V - 127 V AC) is used, the power supply is limited to 600 W.

There is one common power domain for the chassis that distributes DC power to each of the nodes and modules through the system midplane.

DC redundancy is achieved when there is one more power supply available than is needed to provide full power to all chassis components. AC redundancy is achieved by distributing the AC power cord connections between independent AC circuits. For more information, see 5.1, “Power planning” on page 64.

Each power supply includes presence circuitry, which is powered by the midplane. This circuitry allows the FPC to recognize when power supplies are installed in the enclosure but are not powered by the AC supply.

The power supplies support oversubscription. By using oversubscription, users can make the most of the extra power from the redundant power supplies when the power supplies are in a healthy condition. You can use the power capacity of all installed power supplies while still preserving power supply redundancy if there is a power supply failure. For more information, see 3.9.4, “Power supply oversubscription” on page 35.

As shown in Figure 3-5 on page 45, the following LEDs are on each power supply:

- ▶ AC power LED
When this LED is lit (green), it indicates that AC power is supplied to the power supply.
- ▶ DC power LED
When this LED is lit (green), it indicates that DC power is supplied from the power supply to the chassis midplane.
- ▶ Fault LED
When this LED is lit (yellow), it indicates that there is a fault with the power supply.

Removing a power supply: To maintain proper system cooling, do not operate the NeXtScale n1200 Enclosure without a power supply (or power supply filler) in every power supply bay. Install a power supply within 1 minute of the removal of a power supply.

3.5 Fan modules

The NeXtScale n1200 Enclosure supports 10 80 mm fan modules. All fans modules are at the rear of the chassis and are numbered, as shown in Figure 3-7.

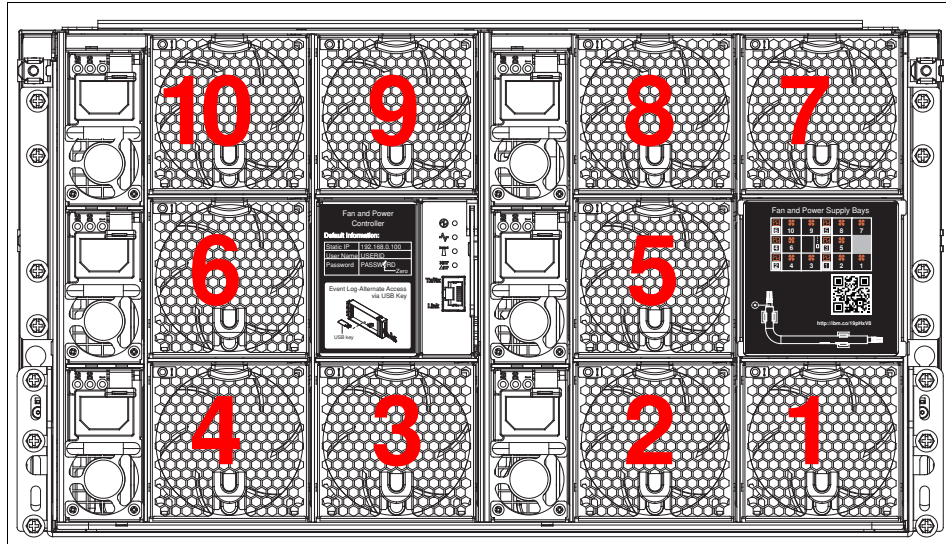


Figure 3-7 NeXtScale n1200 Enclosure rear view with fan bay numbering

The fan modules provide cooling to the compute nodes. The fan modules have a dual-rotor design for high efficiency and high reliability; air flow is front-to-back.

Ordering information for the fan modules is listed in Table 3-32.

Table 3-32 Fan Modules

Part number	Feature code	Description	Maximum supported	Chassis model where used
00Y8570	A41F	n1200 Fan Module	10	A2x

The 80 mm fan module is shown in Figure 3-8.



Figure 3-8 80 mm fan module

Fan module controls and indicators

Each fan module features a Fault LED. When this LED is lit (yellow), it indicates that the fan module failed.

The fan modules are not dedicated to cool a specific node. If there is a fan module failure, the remaining functional fans speed up (if required) under the control of the FPC to provide sufficient cooling to the chassis elements.

There are two logical cooling zones in the enclosure, as shown in Figure 3-9. Five fan modules on each side correspond to the six compute nodes on that same side of the chassis.

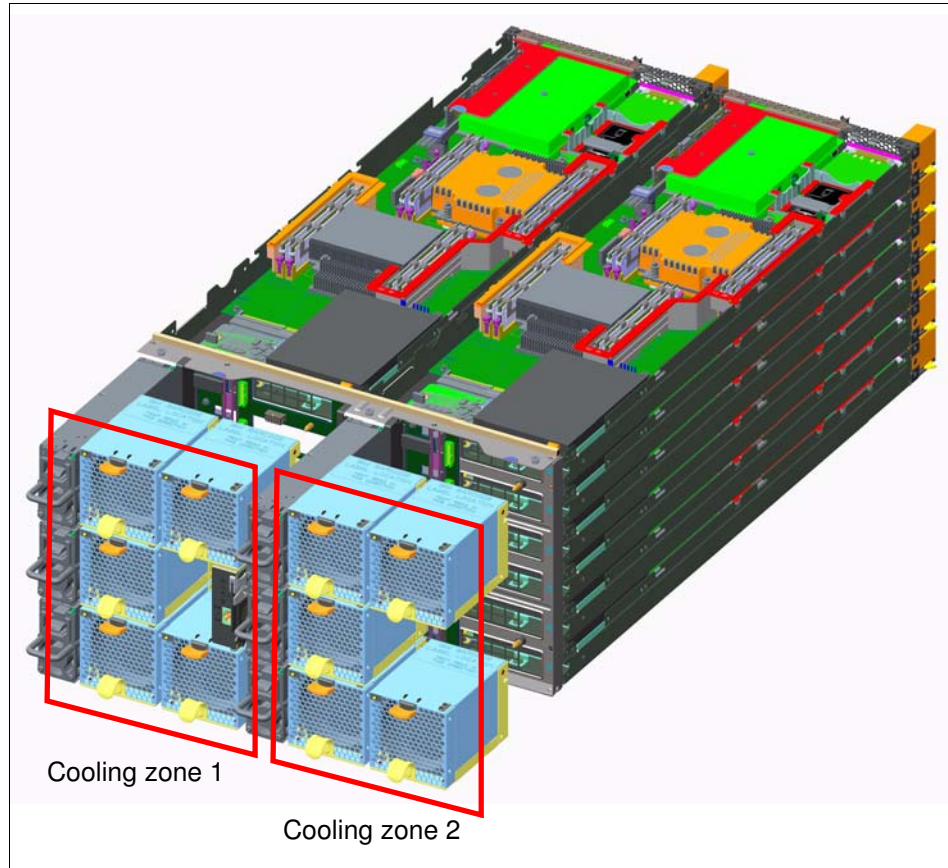


Figure 3-9 Cooling zones on the chassis

For each of the cooling zones (zone 1 or zone 2), the FPC sets the respective fans for the corresponding nodes to the appropriate cooling values that are required to cool those nodes. The FPC varies the speeds of the fans in zone 1 and zone 2 by at most a 20% difference to avoid unbalanced air flow distribution.

Fan removal: To maintain proper system cooling, do not operate the NeXtScale n1200 Enclosure without a fan module (or fan module filler) in every fan module bay. Install a fan module within 1 minute of the removal of a fan module.

3.6 Midplane

The enclosure midplane is the bridge to connect the compute nodes with the power supplies, fan modules, and the FPC. Figure 3-6 shows the front and rear of the midplane.

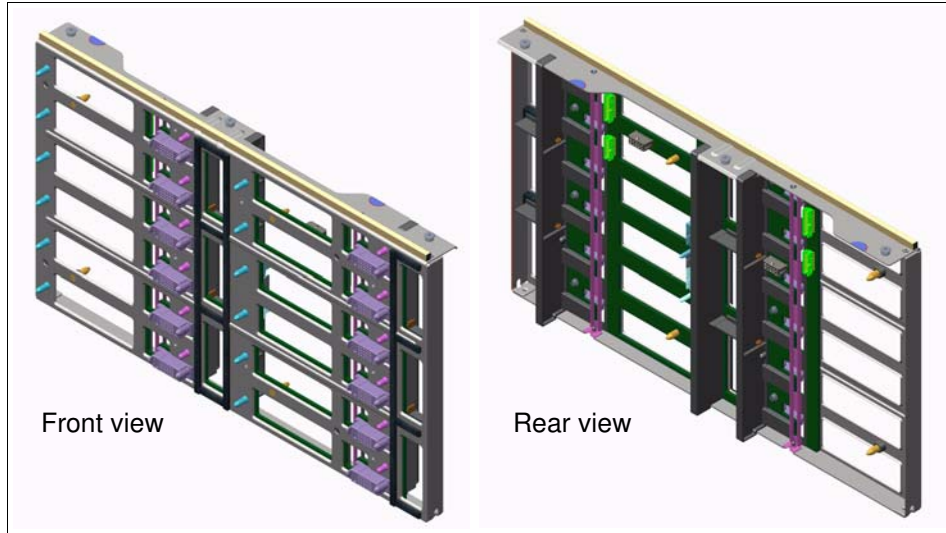


Figure 3-10 Front and rear view of the n1200 Enclosure Midplane Assembly

The midplane is used to provide power to all elements in the chassis. It also provides signals to control fan speed, power consumption, and node throttling.

The midplane was designed with no active components to improve reliability and minimize serviceability. Unlike BladeCenter, the midplane is removed by removing a cover from the top of the chassis.

Figure 3-7 shows the connectivity of the chassis components through the midplane.

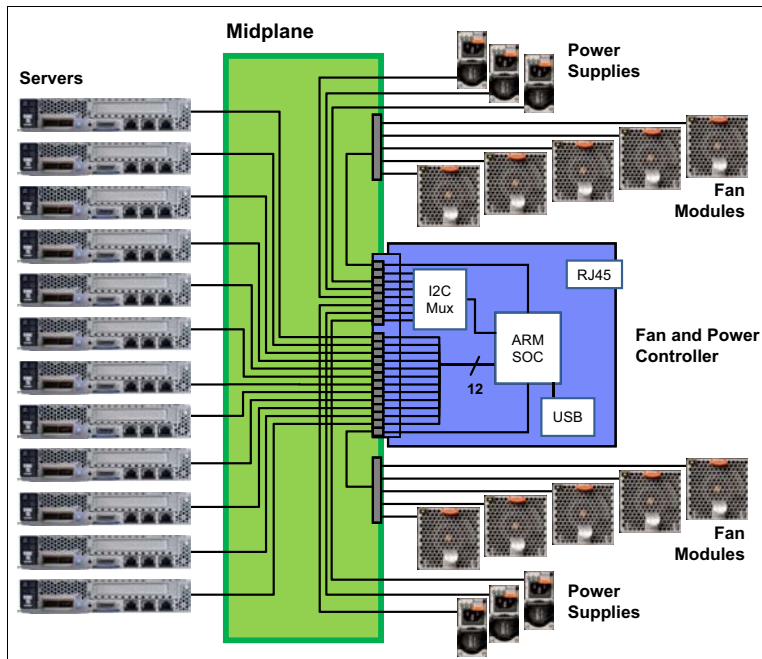


Figure 3-11 Midplane connectivity

3.7 Fan and Power Controller

The Fan and Power Controller (FPC) controls the power budget, provides the power permission to each node, and controls the speed of the fans. The FPC is installed inside the chassis and is accessible from the rear of the chassis, as shown in Figure 3-12. The FPC is a hot-swappable component, as indicated by the orange handle.



Figure 3-12 Rear view of the chassis that shows the location of the FPC

3.7.1 Ports and connectors

The FPC provides integrated systems management functions. The user interfaces (browser and CLI) are accessible remotely via the 10/100 Mbps Ethernet port.

Figure 3-9 shows the FPC and its LEDs.

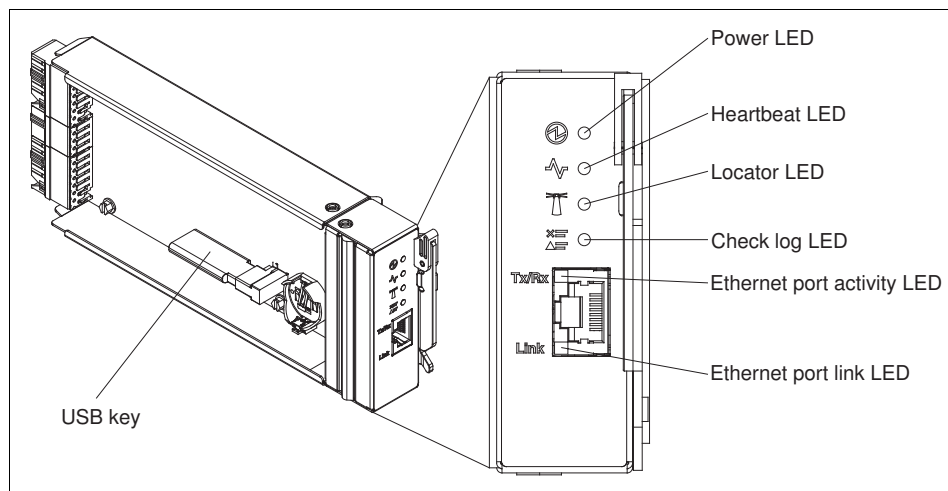


Figure 3-13 FPC

The FPC has the following LEDs and connector that you can use to obtain status information and restart the FPC:

- ▶ Power LED

When this LED is lit (green), it indicates that the FPC has power.

- ▶ Heartbeat LED

When this LED is lit (green), it indicates that the FPC is actively controlling the chassis.

- ▶ Locator LED

When this LED is lit or flashing (blue), it indicates the chassis location in a rack. The locator LED is lights or flashes in response to a request for activation via the FPC web interface or a systems management application.

- ▶ Check log LED

When this LED is lit (yellow), it indicates that a system error occurred.

- ▶ Ethernet port activity LED

When this LED is flashing (green), it indicates that there is activity through the remote management and console (Ethernet) port over the management network.

- ▶ Ethernet port link LED

When this LED is lit (green), it indicates that there is an active connection through the remote management and console (Ethernet) port to the management network.

- ▶ Remote management and console (Ethernet) connector

The remote management and console RJ45 connector is the management network connector for all chassis components. This 10/100 Mbps Ethernet connector is connected to the management network through a top-of-rack switch.

Note: The FPC is not a point of failure for the chassis. If the FPC fails, the compute nodes, power supplies, and fans remain functional to keep the systems running. The power capping policies that are set for the chassis and compute nodes remain in place. The fans speed up to maximum to provide cooling for the compute nodes until the FPC is replaced.

3.7.2 Internal USB memory key

The FPC also includes a USB key that is housed inside the unit, as shown in Figure 3-10.

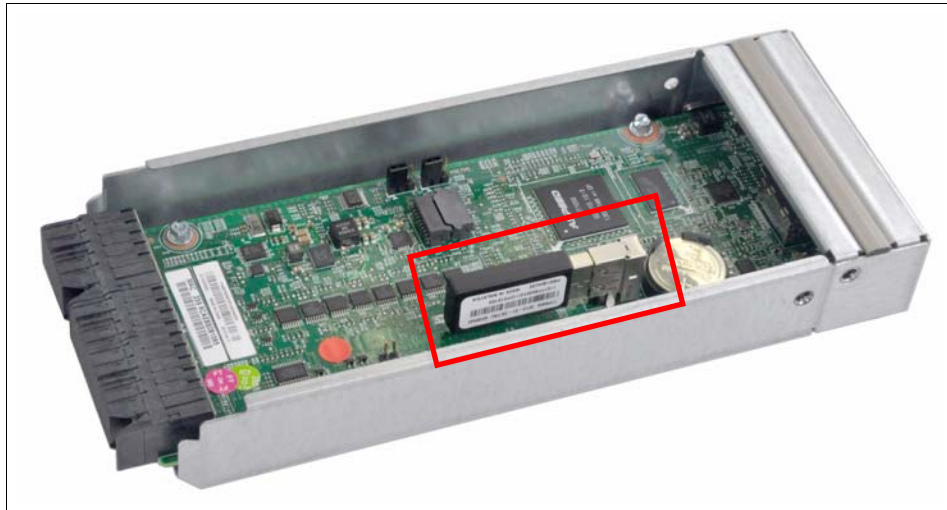


Figure 3-14 Internal view of the FPC

The USB key saves the following information:

- ▶ Event log
- ▶ Enclosure configuration data:
 - PSU redundancy setting
 - Oversubscription mode setting
 - Chassis/node-level power capping value and settings
 - Power restore policy
 - Acoustic mode setting
- ▶ Midplane vital product data (VPD)

If the FPC fails and must be replaced, users can restore the configuration data to the new FPC by transferring the USB from the old unit to the new unit.

3.7.3 Overview of functions

The FPC performs the following functions:

- ▶ Controls the power and power management features of the chassis, compute nodes, and power supplies. The FPC prevents a compute node from powering on if there is insufficient power available from the power supplies.
- ▶ Controls the cooling of the chassis. The FPC ramps up fan speeds if conditions require more cooling or slows down the fans to conserve energy if less cooling is possible.
- ▶ Provides the following user interfaces:
 - Web interface
 - IPMI command line (for external management tools, such as ipmitool or xCAT)
- ▶ Allows you to select one of the following power supply redundancy policies:
 - N+1, where one power supply is redundant and allows for a single power supply to fail without any loss of function.

- N+N, where half the power supplies are redundant backups of the other half. This interface is useful if you have two power utility sources and want the chassis to survive the failure of one of the utility sources.
- No redundancy, which maximizes the power that is available to the compute nodes at the expense of power supply redundancy.
- Oversubscription, which can be enabled with N+1 and N+N policies.
- Smart Redundancy mode, which can disable power conversion on some power supplies to increase the efficiency of the other power supplies during times with low-power requirements.
- ▶ Supports updating the FPC firmware.
- ▶ Monitors and reports fan, power supply, and chassis status and other failures in the event log and with corresponding LEDs.

3.7.4 Web GUI interface

Through the FPC web interface, the user or system administrator can perform the following tasks. For more information about the FPC web interface, see 7.2.1, “FPC web browser interface” on page 123:

- ▶ View summary of elements status:
 - Front and rear view of the chassis
 - Compute nodes location and status
 - FPC, power supplies, fan modules location, and status
- ▶ View current power usage:
 - Voltage overview of the chassis
 - Total chassis power consumption (AC-in)
 - Total PSU power consumption (DC-out)
 - Total fans power consumption
 - Per-node power consumption
 - Power supply fan speeds
- ▶ View and set power supply redundancy, oversubscription, and smart redundancy:
 - Select No Redundancy, N+1, or N+N Redundant mode
 - Enable or disable oversubscription mode
 - Select Disabled or 10-, 30-, or 60-minute scanning periods for Smart Redundancy mode
- ▶ View and set power capping:
 - Node level: Set value within a defined range for each node separately, or choose between one of the three predefined modes.
 - Chassis level: Set value within a defined range for the enclosure, or choose between one of the three predefined modes.
- ▶ View and set power restore policy: Enable or disable (for each node or chassis)
- ▶ View current fan speeds
- ▶ View and set Acoustic mode (three modes)
- ▶ View Chassis, Midplane, and FPC vital product data (VPD) details
- ▶ View, save, and clear the system event log

- ▶ View and set network configuration:
 - SMTP configuration
 - SNMP traps and email alert configuration
 - Host name, DNS, Domain, IP, and IP version configuration
 - SNMP traps email alert configuration
 - Web server (http or https) ports configuration
- ▶ Perform a Virtual Reset or Virtual Reseat of each compute node
- ▶ Set Locator (Identify) LED to on, off, or flash
- ▶ Turn off Check Log LED
- ▶ Back up and restore FPC configuration to USB key; reset to default
- ▶ Perform firmware update
- ▶ Set date and time
- ▶ Perform user account management

3.8 Power management

The FPC controls the power on the NeXtScale n1200 Enclosure. If there is sufficient power available, the FPC allows a compute node to be powered on.

The power permission includes the following two-step process:

1. Pre-boot (BMC stage) inventory power (standby power) is pre-determined based on the node type.
2. Post-boot (UEFI stage) inventory power is a more accurate estimation of the node's maximum power usage that is based on power maximizer test. The following values are generated:
 - Maximum power usage value under stressed condition.
 - Maximum power usage value under stressed condition when P-state is capped at the lowest level.

The FPC uses these values to compare the total node power and total available power to determine power-on and boot permissions.

3.8.1 Power Restore policy

By using the Power Restore policy, you can specify whether you want the compute nodes to restart when a chassis AC power is removed and restored. This policy is similar to the Automatic Server Restart (ASR) feature of many System x servers. For more information about the Power Restore Policy, see “Power Restore Policy tab” on page 132.

When Power Restore policy is enabled, the FPC turns the compute node back on automatically when power is restored if the compute node was powered on before the AC power cycle.

However, a compute node is restarted only if the FPC determines there is sufficient power available to power on the server, which is based on the following factors:

- ▶ Number of working power supplies
- ▶ Power policy that is enabled
- ▶ The oversubscription policy

If there is insufficient power, some compute nodes are not powered on. The nodes are powered on based on their power usage, lowest first. The objective is to maximize the number of nodes that can be powered on in a chassis. The power-on sequence nominally takes approximately 2 minutes with most of that time spent by the nodes running a power maximizer. After that process is complete, the FPC can quickly release the nodes to continue to boot.

3.8.2 Power capping

Users can choose chassis-level capping or saving, or node-level capping or saving, through the power capping configuration options. Power capping allows users to set a wattage limit on power usage. When it is applied to an individual node, the node power consumption is capped at an assigned level. When it is applied to a chassis, the whole chassis power usage is capped. When power saving is enabled, an individual node or all nodes (chassis level) run in modes of different throttling level, depending on the modes that are chosen.

3.8.3 Power supply redundancy modes

The FPC offers the following power supply redundancy modes:

- No redundancy mode

The loss of any power supply can affect the system's operation or performance. If the chassis is evaluated to be vulnerable, because of the failure of one or multiple power supplies, throttle signals are sent to all nodes in the chassis to be throttled down to the lowest power level possible (CPU or Memory lowest P-state). If the power usage remains too high, the chassis is shut down.

This mode is the default mode and does not support the oversubscription mode (see 3.9.4, "Power supply oversubscription" on page 35).

- N+1 Mode

One installed power supply is used as redundant. The failure of one power supply is allowed without affecting the system's operation or performance (performance can be affected if oversubscription mode is enabled).

This mode can be enabled with oversubscription mode.

- N+N Mode

Half of the power supplies that are installed are used as redundant. The failure of up to half the number of the power supplies is allowed without affecting the system's operation or performance (performance can be affected if oversubscription mode is enabled). This mode is useful if you have two power sources from two separate PDU for example.

This mode can be enabled with oversubscription mode.

3.8.4 Power supply oversubscription

By using oversubscription, users can make the most of the extra power from the redundant power supplies when the power supplies are in healthy condition.

For example, when oversubscription is enabled with N+1 redundancy mode, the total power that is available is equivalent to No Redundancy mode with six power supplies in the chassis. This configuration means that the power of six power supplies can be counted on instead of five for normal operation.

When oversubscription mode is enabled with redundant power (N+1 or N+N redundancy), the total available power is 120% of the label ratings of the power supplies. Therefore, for a 1300 W power supply, it can be oversubscribed to $1300\text{ W} \times 120\% = 1,560\text{ W}$.

For example, with an N+1 power policy and six power supplies, instead of $5 \times 1300\text{ W}$ (6500 W) of power, there are $5 \times 1300\text{ W} \times 120\%$ (7800 W) of power that is available to the compute nodes.

Table 3-6 lists the power budget that is available, depending on the redundancy and oversubscription mode that is selected.

Table 3-33 Power budget for 6 x 1300 W power supplies

Redundancy Mode	Oversubscription mode	Power budget ^a
Non-redundant	Not available	7800 W (= 6 x 1300 W)
N+1	Disabled	6500 W (= 5 x 1300 W)
	Enabled	7800 W (= 5 x 1300 W x 120%)
N+N	Disabled	3900 W (= 3 x 1300 W)
	Enabled	4680 W (= 3 x 1300 x 120%)

a. The power budget that is listed in this table is based on power supply ratings. Actual power budget can vary.

When oversubscription mode is enabled with redundant power (N+1 or N+N redundancy), the chassis' total available power can be stretched beyond the label rating (up to 120%). However, the power supplies can sustain this oversubscription for a limited time (approximately 1 second).

In healthy condition (all power supplies are in normal-operational mode), the redundant power supplies provide the extra 20% power oversubscription load for the rest of the normal-operational power supplies (none of the power supplies are oversubscribed).

When redundant power supplies fail (that is, one power supply failure in N+1 mode, or up to N power supplies fail in N+N mode), the remaining normal-operational power supplies provide the extra 20% power oversubscription load. This extra power is provided for a limited time only to allow the compute nodes to throttle to the lowest P-state to reduce their power usage back to a supported range. By design, the compute nodes perform this action quickly enough and operation continues.

Non-redundant mode: It is not possible to enable the oversubscription mode without any power redundancy.

The Table 3-7 on page 37 lists the consequences of redundancy failure in the chassis with and without oversubscription mode.

Table 3-34 Consequences of power supply failure, depending on the oversubscription

Redundancy mode	Oversubscription mode	Consequences of redundancy failure ^a	
		Compute nodes might be throttled ^b	Chassis might power off
Non-redundant	Not available	Yes	Yes ^c
N+1	Disabled	No	No
	Enabled	Yes	No
N+N	Disabled	No	No
	Enabled	Yes	No

- a. Considering one power supply failure in non-redundant and N+1 mode and three power supplies failures in N+N mode.
- b. Compute nodes are throttled only if they require more power than what is available on the remaining power supplies.
- c. The chassis is powered off only if after throttling the compute nodes the enclosure power requirement still exceeds the power that is available on the remaining power supplies.

3.8.5 Acoustic mode

By using acoustic mode, the user has some control over the fan speeds and thus noise that is produced by the system fans. This mode can be used for noise concerns in the user environment. Installation sound pressure levels are also influenced by the number of racks in the installation; the size, materials, and configuration of the room; the noise levels from other equipment; the room ambient temperature and pressure, and measurement location in relation to the equipment.

When the option is set to Off, the power supply fan speeds change as required for optimal cooling.

When the option is set to On, the chassis offers the following set points where the fan speeds are capped:

- ▶ Mode 1: Highest acoustics attenuation (lowest cooling), system power supply fan speeds are capped at 28% duty:
 - 6.93 bels maximum with 900 W power supplies
 - 7.19 bels maximum with 1300 W power supplies
- ▶ Mode 2: Intermediate acoustics attenuation, system power supply fan speeds are capped at 34% duty:
 - 7.27 bels maximum with 900 W power supplies
 - 7.42 bels maximum with 1300 W power supplies
- ▶ Mode 3: Low acoustics attenuation (higher cooling), system power supply fan speeds are capped at 40% duty:
 - 7.58 bels maximum with 900 W power supplies
 - 7.89 bels maximum with 1300 W power supplies

These settings increase the possibility that nodes must be throttled to maintain cooling within the fan speed limitation. If acoustic mode is enabled when the ambient temperature is above 27 °C (81 °F) indefinitely, it is possible that nodes might need to be shut down to prevent overheating. Acoustic mode is automatically disabled if there is a thermally demanding PCI card that is installed in a PCI Expansion Tray on a NeXtScale node in the chassis.

3.8.6 Smart Redundancy mode

Smart Redundancy mode improves power supply efficiency by powering off power supplies that are not needed for the current workload of the chassis and current power supply redundancy settings. The remaining active power supplies operate at a higher load, which improves their efficiency.

However, disabling Smart Redundancy mode always keeps all power supplies active.

The following scanning periods are available:

- ▶ 10 minutes
- ▶ 30 minutes (default)
- ▶ 60 minutes

The shorter the scanning period, the faster the FPC shuts off the DC output section of the selected power supplies.

Smart redundancy is only available with 1300 W power supplies.

Three-phase power: In data centers with three-phase power, the use of Smart Redundancy mode can unbalance the load on the phases, which can lead to larger neutral currents that have the potential to negate the power savings at the chassis level.

3.9 Specifications

This section describes the specifications of the NeXtScale n1200 Enclosure.

3.9.1 Physical specifications

The enclosure features the following physical specifications:

- ▶ Dimensions:
 - Height: 263 mm (10.37 in.)
 - Depth: 915 mm (36 in.)
 - Width: 447 mm (17.6 in.)
- ▶ Weight:
 - Fully configured (stand-alone): Approximately 112 kg (247 lb.)
 - Empty chassis: Approximately 28 kg (62 lb.)
- ▶ Approximate heat output:
 - Ship configuration: 341.18 Btu/hr (100 watts)
 - Full configuration: 20,470.84 Btu/hr (6,000 watts)
- ▶ Declared sound power level: 7.5 bels
- ▶ Chassis airflow:

Full chassis configuration with all compute nodes, FPC, power supplies, and fan modules installed:

 - Minimum: 158 CFM (idle)
 - Maximum: 614 CFM

3.9.2 Supported environment

The NeXtScale n1200 Enclosure complies with the following ASHRAE class A3 specifications.

- ▶ Power on¹:
 - Temperature: 5 °C - 40 °C (41 °F - 104 °F)²
 - Humidity, non-condensing: -12 °C dew point (10.4 °F) and 8% - 85% relative humidity
 - Maximum dew point: 24 °C (75 °F)
 - Maximum altitude: 3048 m (10,000 ft.)
 - Maximum rate of temperature change: 5 °C/hr. (41 °F/hr.)³
- ▶ Power off⁴:
 - Temperature: 5 °C - 45 °C (41 °F - 113 °F)
 - Relative humidity: 8% - 85%
 - Maximum dew point: 27 °C (80.6 °F)
- ▶ Storage (non-operating):
 - Temperature: 1 °C to 60 °C (33.8 °F - 140 °F)
 - Altitude: 3050 m (10,006 ft.)
 - Relative humidity: 5% - 80%
 - Maximum dew point: 29 °C (84.2 °F)
- ▶ Shipment (non-operating)⁵:
 - Temperature: -40°C - 60°C (-40°F - 140°F)
 - Altitude: 10700 m (35,105 ft.)
 - Relative humidity: 5% - 100%
 - Maximum dew point: 29 °C (84.2 °F)⁶

¹ Chassis is powered on.

² A3: Derate maximum allowable temperature 1 °C/175 m above 950 m.

³ 5 °C per hour for data centers that use tape drives and 20 °C per hour for data centers that use disk drives.

⁴ Chassis is removed from original shipping container and is installed but not in use; for example, during repair, maintenance, or upgrade.

⁵ The equipment acclimation period is 1 hour per 20 °C of temperature change from the shipping environment to the operating environment.

⁶ Condensation is acceptable, but not rain.

Compute nodes

NeXtScale System is the new generation dense platform from Lenovo. There are two compute nodes that are supported in the NeXtScale n1200 Enclosure: the nx360 M5 with Intel Xeon E5 v3 processors and the nx360 M4 with Intel Xeon E5 v2 processors.

This chapter includes the following topics:

- ▶ 4.1, “NeXtScale nx360 M5 compute node” on page 62
- ▶ 4.2, “NeXtScale nx360 M4 compute node” on page 110

4.1 NeXtScale nx360 M5 compute node

The NeXtScale nx360 M5 compute node (machine type 5465) is a half-wide, dual-socket server that is designed for data centers that require high performance but are constrained by floor space. It supports Intel Xeon E5-2600 v3 series processors up to 18 cores, which provide more performance per server than previous generation systems.

With more computing power per watt and the latest Intel Xeon processors, you can reduce costs while maintaining speed and availability. A total of 12 nx360 M5 servers can be installed into the 6U NeXtScale n1200 enclosure.

This section describes the nx360 M5 compute node and includes the following topics:

- ▶ 4.1.1, "Overview" on page 63
- ▶ 4.1.2, "System architecture" on page 67
- ▶ 4.1.3, "Standard specifications" on page 69
- ▶ 4.1.4, "Standard models" on page 72
- ▶ 4.1.5, "Processor options" on page 73
- ▶ 4.1.6, "Memory options" on page 74
- ▶ 4.1.7, "NeXtScale 12G Storage Native Expansion Tray" on page 79
- ▶ 4.1.8, "Internal storage" on page 80
- ▶ 4.1.9, "Controllers for internal storage" on page 83
- ▶ 4.1.10, "Internal drive options" on page 88
- ▶ 4.1.11, "I/O expansion options" on page 92
- ▶ 4.1.12, "Network adapters" on page 93
- ▶ 4.1.13, "Storage host bus adapters" on page 95
- ▶ 4.1.14, "NeXtScale PCIe Native Expansion Tray" on page 96
- ▶ 4.1.15, "NeXtScale PCIe 2U Native Expansion Tray" on page 98
- ▶ 4.1.16, "GPU and coprocessor adapters" on page 100
- ▶ 4.1.17, "Integrated virtualization" on page 102
- ▶ 4.1.18, "Local server management" on page 103
- ▶ 4.1.19, "Remote server management" on page 104
- ▶ 4.1.20, "Supported operating systems" on page 106
- ▶ 4.1.21, "Physical and environmental specifications" on page 107
- ▶ 4.1.22, "Regulatory compliance" on page 108

4.1.1 Overview

Figure 4-1 shows the NeXtScale nx360 M5 server.

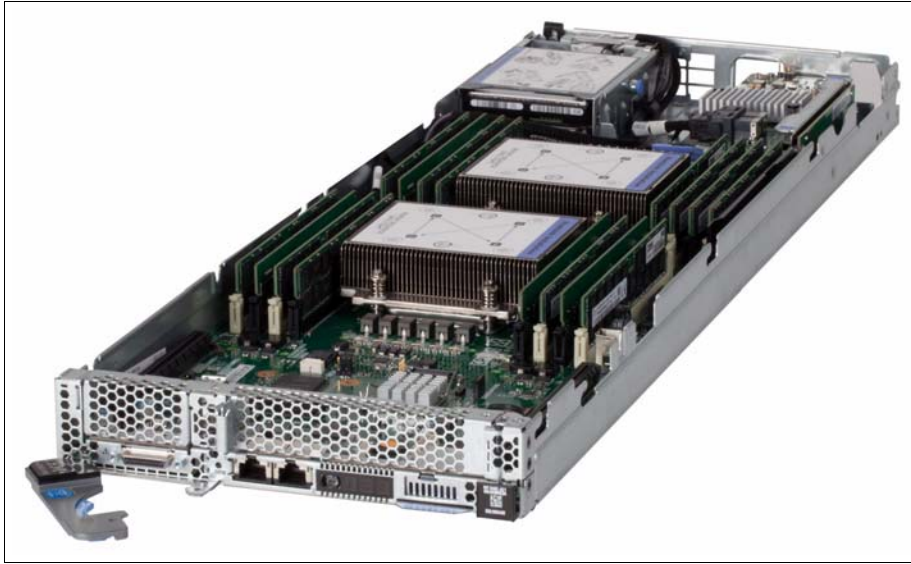


Figure 4-1 NeXtScale nx360 M5 server is based on the Intel Xeon E5-2600 v3 processor family

Scalability and performance

The NeXtScale nx360 M5 server offers the following features to boost performance, improve scalability, and reduce costs:

- ▶ Up to 12 compute nodes, each with two of the latest Xeon processors, 16 DIMMs, and three PCIe slots, in 6U of rack space. It is a highly dense, scalable, and price-optimized offering.
- ▶ The Intel Xeon processor E5-2600 v3 product family improves productivity by offering superior system performance with 18-core processors, core speeds up to 3.2 GHz, L3 cache sizes up to 45 MB, DDR4 memory speeds up to 2133 MHz, and QPI interconnect links of up to 9.6 GTps.
- ▶ Two processors, up to 36 cores, and 72 threads maximize the concurrent running of multi-threaded applications.
- ▶ Intelligent and adaptive system performance with Intel Turbo Boost Technology 2.0 allows CPU cores to run at maximum speeds during peak workloads by temporarily going beyond processor thermal design power (TDP).
- ▶ Intel Hyper-Threading Technology boosts performance for multi-threaded applications by enabling simultaneous multi-threading within each processor core, up to two threads per core.
- ▶ Intel Virtualization Technology integrates hardware-level virtualization hooks that allow operating system vendors to better use the hardware for virtualization workloads.
- ▶ Intel Advanced Vector Extensions 2 (AVX2) improve floating-point performance for compute-intensive technical and scientific applications.
- ▶ Sixteen DIMMs of registered 2133 MHz DDR4 ECC memory provide speed, high availability, and a memory capacity of up to 1 TB.
- ▶ Supports drives up to 6 TB capacity in the 3.5-inch form factor.
- ▶ Support for internal simple-swap drives: one 3.5-inch drive, two 2.5-inch drives, or four 1.8-inch drives. Also, two 2.5-inch hot-swap drives can be added in place of a PCIe slot.

- ▶ Support for more local storage with the use of the 12G Storage Native Expansion Tray. When 6 TB hard disk drives (HDDs) are used, you can create an ultra-dense storage server with up to 48 TB of total disk capacity within 1U of comparable rack density. The nx360 M5 with the Storage Native Expansion Tray offers a perfect solution for today's data-intensive workloads.
- ▶ Boosts performance with PCIe Native Expansion Trays by offering support for up to four high-powered GPUs or Intel Xeon Phi coprocessors within a single node.
- ▶ The use of solid-state drives (SSDs) instead of or with traditional HDDs can improve I/O performance. An SSD can support up to 100 times more I/O operations per second (IOPS) than a typical HDD.
- ▶ Three PCIe slots internal to the nx360 M5: Full-height PCIe slot, mezzanine LOM Generation 2 (ML2) slot, and dedicated internal RAID adapter slot.
- ▶ Supports new ML2 cards for 40 Gb Ethernet and FDR InfiniBand that offer network performance in the smallest footprint.
- ▶ PCI Express 3.0 I/O expansion capabilities improve the theoretical maximum bandwidth by 60% compared with the previous generation of PCI Express 2.0.
- ▶ With Intel Integrated I/O Technology, the PCI Express 3.0 controller is integrated into the Intel Xeon processor E5 family, which reduces I/O latency and increases overall system performance.

Manageability and security

The following powerful systems management features simplify local and remote management of the nx360 M5:

- ▶ The server includes an Integrated Management Module II (IMM 2.1) to monitor server availability and perform remote management.
- ▶ There is a standard Ethernet port that can be shared between the operating system and IMM for remote management with optional Features on Demand upgrade. There is an optional extra Ethernet port for dedicated IMM connectivity.
- ▶ An integrated industry-standard Unified Extensible Firmware Interface (UEFI) enables improved setup, configuration, and updates that also simplifies error handling.
- ▶ Integrated Trusted Platform Module (TPM) 1.2 support enables advanced cryptographic functions, such as digital signatures and remote attestation.
- ▶ Intel Trusted Execution Technology provides enhanced security through hardware-based resistance to malicious software attacks, which allows the application to run in its own isolated space that is protected from all other software that is running on a system.
- ▶ The Intel Execute Disable Bit function can prevent certain classes of malicious buffer overflow attacks when combined with a supporting operating system.

Energy efficiency

The NeXtScale System offers the following energy efficiency features to save energy, reduce operational costs, increase energy availability, and contribute to a green environment:

- ▶ Support for S3 standby power states in the processor.
- ▶ Shared 80 PLUS Platinum and 80 PLUS Titanium-certified power supplies ensure energy efficiency.
- ▶ Large 80 mm fans maximize air flow efficiencies.
- ▶ The Intel Xeon processor E5-2600 v3 product family offers better performance per watt over the previous generation.

- ▶ Intel Intelligent Power Capability powers on and off individual processor elements as needed to reduce power draw.
- ▶ Low-voltage Intel Xeon processors draw less energy to satisfy the demands of power and thermally constrained data centers and telecommunication environments.
- ▶ Low-voltage 1.2 V DDR4 memory DIMMs use up to 20% less energy compared to 1.35 V DDR3 DIMMs.
- ▶ SSDs use as much as 80% less power than traditional 2.5-inch HDDs.
- ▶ The server uses hexagonal ventilation holes in the front and rear of the casing, which can be grouped more densely than round holes. This feature provides more efficient airflow through the system.
- ▶ There are power monitoring and power capping capabilities through the Power and Fan Management Module in the chassis.

Availability and serviceability

The NeXtScale System and the nx360 M5 server provide the following features to simplify serviceability and increase system uptime:

- ▶ The NeXtScale n1200 chassis supports N+N and N+1 power policies for its six power supplies, which means greater system uptime.
- ▶ All components can be removed from the front of the rack by sliding out the trays or the chassis for easy, quick servicing.
- ▶ Toolless cover removal provides easy access to upgrades and serviceable parts, such as HDDs and memory.
- ▶ The nx360 M5 offers memory mirroring for redundancy if there is a non-correctable memory failure.
- ▶ Optional RAID arrays enable the server to keep operating if there is a failure of any one drive.
- ▶ SSDs offer better reliability than traditional mechanical HDDs for greater uptime.
- ▶ Predictive Failure Analysis (PFA) detects when system components (processors, memory, and HDDs) operate outside of standard thresholds and generates proactive alerts in advance of possible failure, which increases uptime.
- ▶ The built-in IMM2 continuously monitors system parameters, triggers alerts, and performs recovering actions if there are failures to minimize downtime.
- ▶ The IMM2 offers optional remote management capability to enable remote keyboard, video, and mouse (KVM) control of the server.
- ▶ There is a three-year customer replaceable unit and onsite limited warranty, with next business day 9x5. Optional service upgrades are available.

Locations of key components and connectors

Figure 4-2 shows the front of the nx360 M5 server.

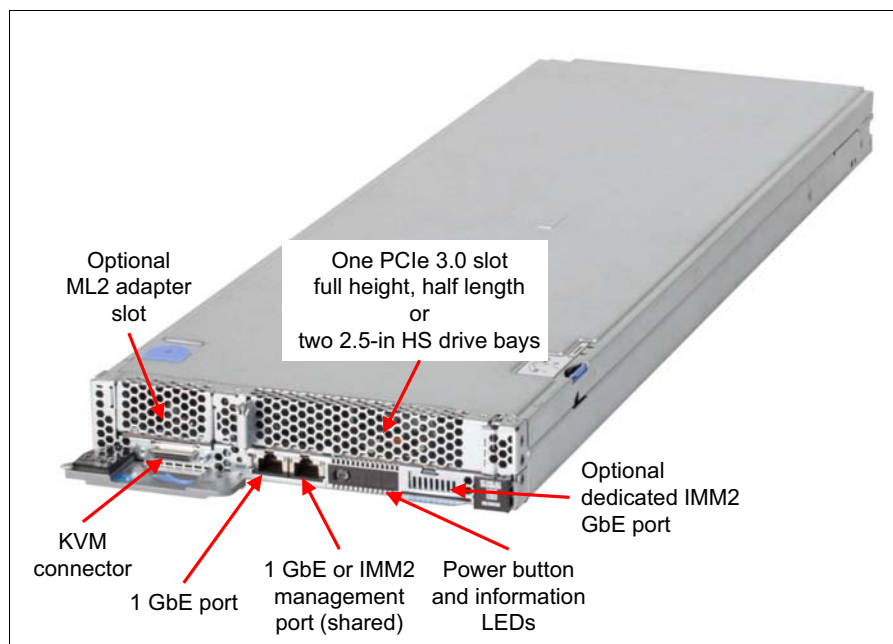


Figure 4-2 Front view of the NeXtScale nx360 M5

Figure 4-3 shows the locations of key components inside the server.

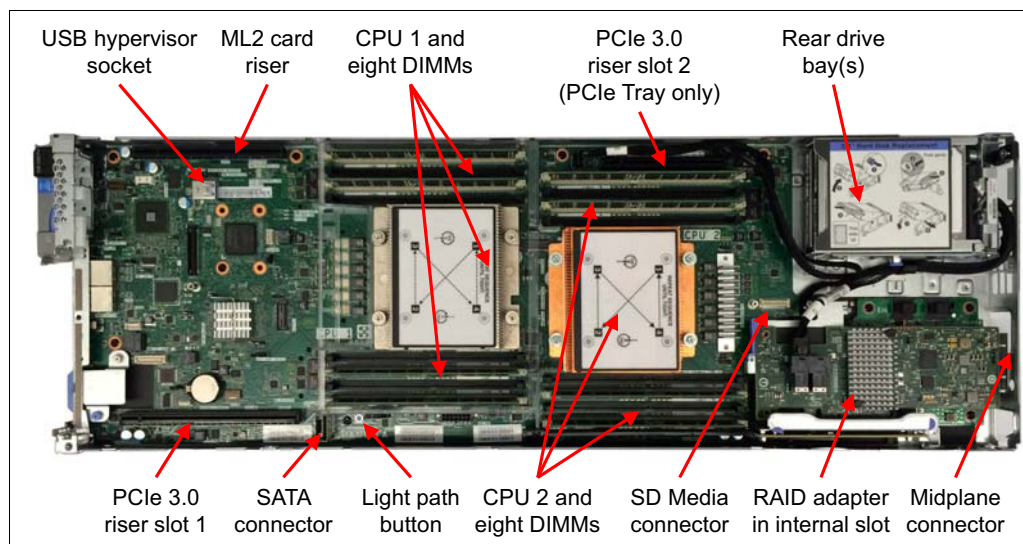


Figure 4-3 Inside view of the NeXtScale nx360 M5

4.1.2 System architecture

The NeXtScale nx360 M5 compute node features the Intel E5-2600 v3 series processors. The Xeon E5-2600 v3 series processor has models with 4, 6, 8, 10, 12, 14, 16, or 18 cores per processor with up to 36 threads per socket.

The Intel Xeon E5-2600 v3 series processor (formerly known by the Intel code name *Haswell-EP*) is the third implementation of Intel's micro architecture that is based on tri-gate transistors. It uses a 22nm manufacturing process.

The processor architecture enables sharing data on-chip through a high-speed ring that is interconnected between all processor cores, the last level cache (LLC), and the system agent. The system agent houses the memory controller and a PCI Express root complex that provides 40 PCIe 3.0 lanes.

The Integrated memory controller in each CPU supports four memory channels with three¹ DDR4 DIMMs per channel that are running at a speed that is up to 2133 MHz. Two QPI links connect the two CPU in a dual-socket installation.

The Xeon E5-2600 v3 series is available with up to 18 cores and 45 MB of last-level cache. It features an enhanced instruction set that is called Intel Advanced Vector Extensions 2 (AVX2). Intel AVX2 extends the Intel AVX with 256-bit integer instructions, floating-point fused multiply add (FMA) instructions and gather operations. Intel AVX2 doubles the number of flops per clock, which doubles the core's theoretical peak floating point throughput. However, when some Intel AVX instructions are run, the processor might run at a less than rated frequency to remain within the TDP limit.

Table 4-1 lists the improvements of the instruction set of the Intel Xeon E5-2600 v3 over previous generations.

Table 4-1 Instructions sets and floating point operations per cycle of Intel processors

Processor Family	Instruction Set	Single Precision Flops Per Clock	Double Precision Flops Per Clock
Intel Xeon 5500 Series (Nehalem)	SSE 4.2	8	4
Intel Xeon E5-2600 and v2 (Sandy Bridge / Ivy Bridge)	AVX	16	8
Intel Xeon E5-2600 v3 (Haswell)	AVX2	32	16

The implementation architecture includes Intel Turbo Boost Technology 2.0 and improved power management capabilities. Intel Turbo Boost Technology dynamically increases the processor's frequency as needed by using thermal and power headroom to provide a burst of speed when the workload needs it and increased energy efficiency when it does not.

As with iDataPlex servers, NeXtScale servers support S3 mode. S3 allows systems to come back into full production from low-power state much quicker than a traditional power-on. Cold boot normally takes approximately 270 seconds; with S3, cold boot occurs in approximately 45 seconds. When you know that a system is not be used because of time of day or state of job flow, you can send it into a low-power state to save power and bring it back online quickly when needed.

¹ Only 2 DIMMs per channel implemented on the nx360 M5 compute node

Table 4-2 lists the differences between the current and the previous generation of Intel's micro architecture implementations (improvements are highlighted in gray).

Table 4-2 Comparison between Xeon E5-2600 v2 and Xeon E5-2600 v3

	Xeon E5-2600 v2 (Ivy Bridge-EP)	Xeon E5-2600 v3 (Haswell-EP)
QPI Speed (GT/s)	8.0, 7.2 and 6.4 GT/s	9.6, 8.0 and 6.4 GTps
Addressability	46 bits physical, 48 bits virtual	
Cores	Up to 12	Up to 18
Threads per socket	Up to 24 threads	Up to 36 threads
Last-level Cache (LLC)	Up to 30 MB	Up to 45MB
Intel Turbo Boost Technology	Yes	
Memory population	4 channels of up to 3 RDIMMs, 3 LRDIMMs, or 2 UDIMMs	4 channels of up to 3 DIMMs per channel and 24 DIMM slots
Maximum memory speed	Up to 1866 MHz DDR3	Up to 2133 MHz DDR4
Memory RAS features	ECC, Patrol Scrubbing, Sparring, Mirroring, Lockstep Mode, x4/x8 SDDC	
PCIe lanes	40 PCIe 3.0 lanes at 8 GTps	40 PCIe 3.0 lanes at 10GTps
TDP values (W)	130, 115, 96, 80, 70, 60, 50 W	165, 145, 135, 120, 105, 90, 85, 65, 55 W
Idle power targets (W)	10.5 W or higher 7.5 W for low-voltage SKUs	9 W or higher 7.5 W for low-voltage SKUs
Instruction Set	AVX	AVX2

Figure 4-4 shows the NeXtScale nx360 M5 system board block diagram.

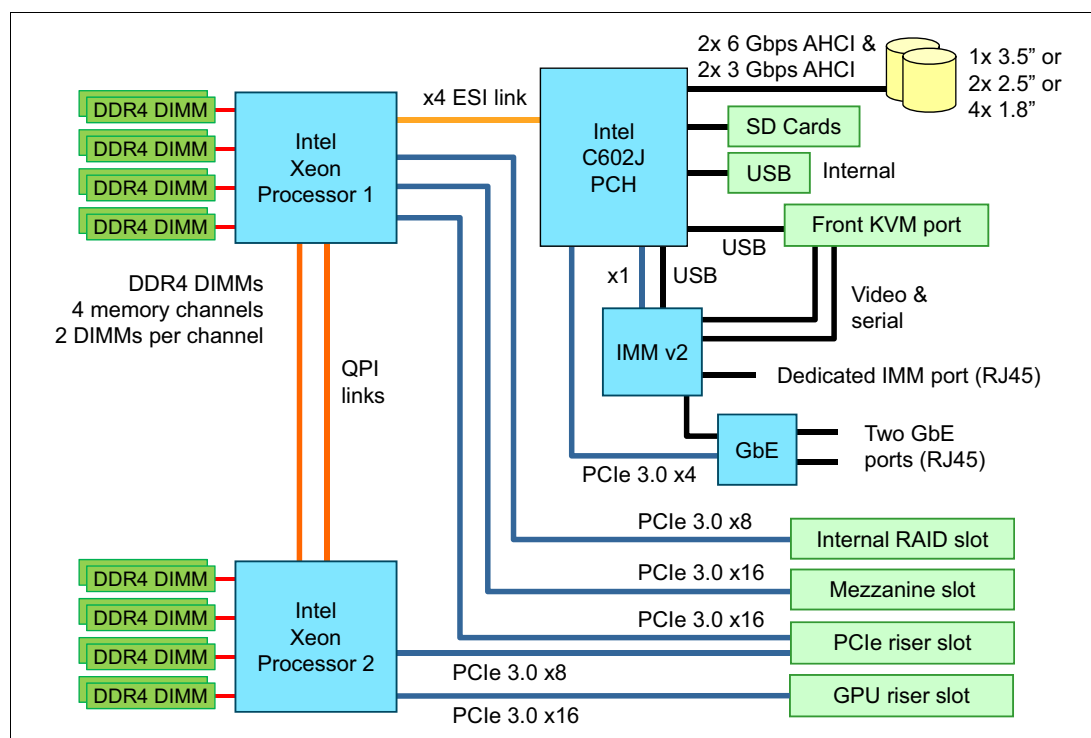


Figure 4-4 NeXtScale nx360 M5 system board block diagram

4.1.3 Standard specifications

Table 4-3 lists the standard specifications of the NeXtScale nx360 M5 compute node.

Table 4-3 Standard specifications

Components	Specification
Machine type	5465
Firmware	Lenovo signed firmware
Form factor	<ul style="list-style-type: none"> Standard server: Half-wide, 1U compute node Optional Native Expansion Tray (PCIe Tray or Storage Tray): Half-wide 2U compute node
Supported chassis	NeXtScale n1200 enclosure, 6U high; up to 12 compute nodes per chassis
Processor	<p>Two Intel Xeon Processor E5-2600 v3 series processors; QuickPath Interconnect (QPI) links speed up to 9.6 GTps. Hyper-Threading Technology and Turbo Boost Technology. Intel C612 chipset. Available core counts:</p> <ul style="list-style-type: none"> 4-core processors at 3.5 GHz and 15 MB L3 cache 6-core processors up to 3.4 GHz and 20 MB L3 cache 8-core processors up to 3.2 GHz and 20 MB L3 cache 10-core processors up to 2.6 GHz and 25 MB L3 cache 12-core processors up to 2.6 GHz and 30 MB L3 cache 14-core processors up to 2.6 GHz and 35 MB L3 cache 16-core processors at 2.3 GHz and 40 MB L3 cache 18-core processors at 2.3 GHz and 45 MB L3 cache

Components	Specification
Memory	Up to 16 DIMM sockets (8 DIMMs per processor) that support DDR4 DIMMs up to 2133 MHz memory speeds. RDIMMs and LRDIMMs also are supported. Four memory channels per processor (two DIMMs per channel).
Memory maximum	LRDIMMs: Up to 1 TB memory with 16x 64 GB LRDIMMs and two processors. RDIMMs: Up to 512 GB memory with 16x 32 GB RDIMMs and two processors.
Memory protection	Chipkill (x4 memory options only), ECC, memory mirroring ^a , and memory rank sparing ^a .
Disk drive bays	<ul style="list-style-type: none"> Internal to the nx360 M5 (not front accessible): One 3.5-inch simple-swap SATA or two 2.5-inch simple swap SAS/SATA HDDs or SSDs, or four 1.8-inch simple-swap SSDs. Front-accessible bays: Two 2.5-inch hot-swap drive bays (optional, replaces the full-height PCIe slot, only supported if internal drive bays are also 2.5-inch bays). With the addition of the NeXtScale 12G Storage Native Expansion Tray (only supported by internal drive bay, not front accessible): Adds seven 3.5-inch simple-swap drive bays.
Maximum internal storage	<p>Without any expansion tray attached:</p> <ul style="list-style-type: none"> With a single 3.5" drive: 6.0 TB using 1x 6TB 3.5" drive (internal) With 2.5" drives: 10.1 TB using 2x 1.2 TB HDDs (internal) + 2x 3.84 TB HS SSDs (front) With 1.8" drives: 1.6 TB using 4x 400 GB SSDs (internal) <p>With the Storage Native Expansion Tray attached:</p> <ul style="list-style-type: none"> All 3.5" drives: 48 TB using 8x 6 TB 3.5" drives (1 internal, 7 in the tray) With 2.5" drives internally: 44.4 TB using 2x 1.2 TB HDDs (internal) + 7x 6 TB 3.5" drives Using 1.8" drives: 43.2 TB using 4x 400 GB SSDs (internal) + 7x 6 TB 3.5" drives <p>With the PCIe 2U Native Expansion Tray attached:</p> <ul style="list-style-type: none"> 3.5" internal drive: 21.3 TB using 1x 6TB 3.5" drive (internal) + 4x 3.84 TB HS SSDs (tray) 2.5" internal drives: 17.8 TB using 2x 1.2 TB HDDs (internal) + 4x 3.84 TB HS SSDs (tray) 1.8" internal drives: 16.9 TB using 4x 400 GB SSDs (internal) + 4x 3.84 TB HS SSDs (tray)
RAID support	<ul style="list-style-type: none"> Four 6 Gb SATA ports through onboard Intel C612 chipset. No RAID standard. Optional 12 Gb SAS/SATA RAID adapters: ServeRAID M5210 or M1215, both standard with RAID 0 and 1. Optional M5210 upgrades: RAID 5, 50 (zero-cache, or 1 GB non-backed cache, or 1 GB or 2 GB or 4 GB flash-backed cache), RAID 6, 60, FoD performance upgrades; optional upgrade to M1215 for RAID 5 support (zero-cache).
Optical drive bays	No internal bays; use an external USB drive. For more information about options, see this website: http://support.lenovo.com/en/documents/pd011281
Tape drive bays	No internal bays. Use an external USB drive.

Components	Specification
Network interfaces	Integrated two-port Gigabit Ethernet (Broadcom BCM5717) with RJ45 connectors. One port that is dedicated for use by the operating system, and one configurable as shared by the operating system and IMM or as dedicated to the IMM. Optional third GbE port for dedicated IMM access. Optionally, PCIe and Mezzanine LOM Gen 2 (ML2) adapters can be added to provide more network interfaces. ML2 Ethernet adapters support shared access to the IMM.
PCI Expansion slots	<p>nx360 M5 without PCIe Native Expansion Tray:</p> <ul style="list-style-type: none"> ▶ One PCIe 3.0 x16 ML2 adapter slot ▶ One PCIe 3.0 x16 full-height half-length slot (or two 2.5" hot-swap drive bays) ▶ One PCIe 2.0 x8 slot for internal RAID controller <p>nx360 M5 with PCIe Native Expansion Tray:</p> <ul style="list-style-type: none"> ▶ One PCIe 3.0 x16 ML2 adapter slot ▶ One PCIe 3.0 x8 full-height half-length slot (or two 2.5" hot-swap drive bays) ▶ One PCIe 2.0 x8 slot for internal RAID controller ▶ Two PCIe 3.0 x16 full-height full-length double-width slots <p>nx360 M5 with 12G Storage Native Expansion Tray:</p> <ul style="list-style-type: none"> ▶ One PCIe 3.0 x16 ML2 adapter slot ▶ One PCIe 3.0 x16 full-height half-length slot for RAID controller <p>nx360 M5 with PCIe 2U Native Expansion Tray:</p> <ul style="list-style-type: none"> ▶ One PCIe 3.0 x16 ML2 adapter slot ▶ One PCIe 3.0 x8 full-height half-length slot ▶ One PCIe 2.0 x8 slot for internal RAID controller ▶ Four PCIe 3.0 x16 full-height full-length double-width slots
Ports	<ul style="list-style-type: none"> ▶ Front of the server: KVM connector; with the addition of a console breakout cable (1 cable standard with the chassis) supplies one RS232 serial port, one VGA port, and two USB 1.1 ports for local console connectivity. Two 1 Gbps Ethernet ports with RJ45 connectors. Optional third GbE port for dedicated IMM2 access. ▶ Internal: One internal USB port for VMware ESXi hypervisor key. Optional support for SD Media Adapter for VMware vSphere hypervisor.
Cooling	Supplied by the NeXtScale n1200 enclosure; 10 hot-swap dual-rotor 80 mm system fans with tool-less design.
Power supply	Supplied by the NeXtScale n1200 enclosure. Up to six hot-swap power supplies 900 W or 1300 W or 1500 W, depending on the chassis model. Support power policies N+N or N+1 power redundancy and non-redundant; 80 PLUS Platinum or Titanium certified depending on the power supply selected.
Systems management	UEFI, Integrated Management Module II (IMM2.1) with Renesas SH7758 controller, Predictive Failure Analysis, Light Path Diagnostics, Automatic Server Restart, and ServerGuide. Browser-based chassis management through an Ethernet port on the Fan and Power Controller at the rear of the n1200 enclosure. IMM2 upgrades are available to IMM2 Standard and IMM2 Advanced for web GUI and remote presence features.
Video	Matrox G200eR2 video core with 16 MB DDR3 video memory that is integrated into the IMM2. Maximum resolution is 1600 x 1200 with 16M colors (32 bpp) at 75 Hz, or 1680 x 1050 with 16M colors at 60 Hz.
Security	Power-on password, administrator's password, and Trusted Platform Module 1.2.

Components	Specification
Operating systems supported	Windows Server 2012 and 2012 R2, SUSE Linux Enterprise Server 11 and 12, Red Hat Enterprise Linux 6 and 7, VMware vSphere 5.1, 5.5 and 6.0.
Limited warranty	Three-year customer-replaceable unit and onsite limited warranty with 9x5/NBD.
Service and support	Optional service upgrades are available through Lenovo Services: 4-hour or 2-hour response time, 8-hour fix time, 1-year or 2-year warranty extension, remote technical support for hardware and some Lenovo and OEM software.
Dimensions	nx360 M5 server: <ul style="list-style-type: none"> ▶ Width: 216 mm (8.5 in.) ▶ Height: 41 mm (1.6 in.) ▶ Depth: 659 mm (25.9 in.).
Weight	nx360 M5 maximum weight: 6.17 kg (13.6 lb).

a. planned for 3Q/2015

The nx360 M5 servers are shipped with the following items:

- ▶ Statement of Limited Warranty
- ▶ Important Notices
- ▶ Documentation CD that contains the Installation and Service Guide

4.1.4 Standard models

Table 4-4 lists the nx360 M5 standard models.

Table 4-4 Standard models

Model	Intel Xeon Processor ^a (2 maximum)	Memory and speed	RAID controller	Drive bays	Disks	Network	Optical
5465-22x	2x E5-2620 v3 6C 2.4GHz 15MB 1866MHz 85W	2x 8 GB 2133 MHz	6 Gbps SATA (No RAID)	1x3.5-inch SS bay	Open	2x GbE	None
5465-42x	2x E5-2650 v3 10C 2.3GHz 25MB 2133MHz 105W	2x 8 GB 2133 MHz	6 Gbps SATA (No RAID)	1x3.5-inch SS bay	Open	2x GbE	None
5465-62x	2x E5-2680 v3 12C 2.5GHz 30MB 2133MHz 120W	2x 16 GB 2133 MHz	6 Gbps SATA (No RAID)	2x2.5-inch SS bays	Open	2x GbE	None

a. Processor detail: Processor quantity and model, cores, core speed, L3 cache, memory speed, and power consumption.

For information about the standard features of the server, see 4.1.3, “Standard specifications” on page 69.

4.1.5 Processor options

The nx360 M5 supports the processor options that are listed in Table 4-5.

Table 4-5 Processor options

Part number	Feature code ^a	Intel Xeon processors ^b	Where used
00FL166	A5HT / A5JA	Intel Xeon Processor E5-2603 v3 6C 1.6GHz 15MB 1600MHz 85W	-
00FL165	A5HS / A5J9	Intel Xeon Processor E5-2609 v3 6C 1.9GHz 15MB 1600MHz 85W	-
00FL163	A5HQ / A5J7	Intel Xeon Processor E5-2620 v3 6C 2.4GHz 15MB 1866MHz 85W	22x
00KA946	AS4N / AS4R	Intel Xeon Processor E5-2623 v3 4C 3.0GHz 10MB 1866MHz 105W	-
00FL162	A5HP / A5J6	Intel Xeon Processor E5-2630 v3 8C 2.4GHz 20MB 1866MHz 85W	-
00FL164	A5HR / A5J8	Intel Xeon Processor E5-2630L v3 8C 1.8GHz 20MB 1866MHz 55W	-
00FL169	A5HW / A5JD	Intel Xeon Processor E5-2637 v3 4C 3.5GHz 15MB 2133MHz 135W	-
00FL161	A5HN / A5J5	Intel Xeon Processor E5-2640 v3 8C 2.6GHz 20MB 1866MHz 90W	-
00FL168	A5HV / A5JC	Intel Xeon Processor E5-2643 v3 6C 3.4GHz 20 MB 2133MHz 135W	-
00FL159	A5HL / A5J3	Intel Xeon Processor E5-2650 v3 10C 2.3GHz 25MB 2133MHz 105W	42x
00FL160	A5HM / A5J4	Intel Xeon Processor E5-2650L v3 12C 1.8GHz 30MB 2133MHz 65W	-
00FL158	A5HK / A5J2	Intel Xeon Processor E5-2660 v3 10C 2.6GHz 25MB 2133MHz 105W	-
00FL167	A5HU / A5JB	Intel Xeon Processor E5-2667 v3 8C 3.2GHz 20MB 2133MHz 135W	-
00FL157	A5HJ / A5J1	Intel Xeon Processor E5-2670 v3 12C 2.3GHz 30MB 2133MHz 120W	-
00FL156	A5HH / A5J0	Intel Xeon Processor E5-2680 v3 12C 2.5GHz 30MB 2133MHz 120W	62x
00KA829	A5V0 / A5V1	Intel Xeon Processor E5-2683 v3 14C 2.0GHz 35MB 2133MHz 120W	-
00KG692	ASGP / ASGQ	Intel Xeon Processor E5-2685 v3 12C 2.6GHz 30MB 2133MHz 120W	-
00FL155	A5HG / A5HZ	Intel Xeon Processor E5-2690 v3 12C 2.6GHz 30MB 2133MHz 135W	-
00FL154	A5HF / A5HY	Intel Xeon Processor E5-2695 v3 14C 2.3GHz 35MB 2133MHz 120W	-
00FL153	A5HE / A5HX	Intel Xeon Processor E5-2697 v3 14C 2.6GHz 35MB 2133MHz 145W	-
00KA945	AS4L / AS4P	Intel Xeon Processor E5-2698 v3 16C 2.3GHz 40MB 2133MHz 135W	-
00KA947	AS4M / AS4Q	Intel Xeon Processor E5-2699 v3 18C 2.3GHz 45MB 2133MHz 145W	-

a. The first feature code corresponds to the first processor; the second feature code corresponds to the second processor.

b. Processor detail: Model, core count, core speed, L3 cache, memory speed, and TDP power.

Floating point performance: The number of sockets and the processor option that are selected determine the theoretical double precision floating point peak performance, as shown in the following example:

$$\text{\#sockets} \times \text{\#cores per processor} \times \text{freq} \times 16 \text{ flops per cycle} = \text{\#Gflops}$$

An nx360 M5 compute node with dual socket E5-2698 v3 series 16-core that operates at 2.3 GHz has the following peak performance:

$$2 \times 16 \times 2.3 \times 16 = 1177.6 \text{ Gflops}$$

However, because of the higher power consumption that is generated by AVX instructions, the processor frequency often is reduced by 0.4 GHz to stay within the TDP limit (The processor frequency drop is not identical for each SKU). As a result, a more realistic peak performance formula is as follows:

$$\text{\#sockets} \times \text{\#cores per processor} \times (\text{freq} - 0.4) \times 16 \text{ flops per cycle} = \text{\#Gflops}$$

As a result, an nx360 M5 compute node with dual socket E5-2698 v3 series 16-core that operates at 2.3 GHz has the following peak performance:

$$2 \times 16 \times (2.3 - 0.4) \times 16 = 972.8 \text{ Gflops}$$

By using LINPACK benchmark, single node sustained performance is approximately 92% of the peak, which corresponds to approximately 895 Gflops with the M5 node that is used in this example.

Memory performance: The processors with eight cores and less have a single memory controller. The processors with 10 cores and more have two memory controllers. As a result, it is recommended to select a 10+ core processor to achieve maximum performance on memory-intensive applications.

4.1.6 Memory options

TruDDR4 Memory uses the highest quality components that are sourced from Tier 1 DRAM suppliers and only memory that meets the strict requirements of Lenovo is selected. It is compatibility tested and tuned on every System x server to maximize performance and reliability.

TruDDR4 Memory has a unique signature that is programmed into the DIMM that enables System x servers to verify whether the memory installed is qualified or supported by Lenovo. Because TruDDR4 Memory is authenticated, certain extended memory performance features can be enabled to extend performance over industry standards. From a service and support standpoint, memory automatically assumes that the Lenovo system warranty and Lenovo provides service and support worldwide.

The NeXtScale nx360 M5 supports up to eight TruDDR4 Memory DIMMs when one processor is installed and up to 16x DIMMs when two processors are installed. Each processor has four memory channels, and there are two DIMMs per memory channel (2 DPC). RDIMMs and LRDIMMs are supported, but the mixing of these different types is not supported.

Table 4-6 on page 75 lists the memory options that are available for the nx360 M5 server.

Table 4-6 Memory options

Part number	Feature code	Description	Maximum supported	Models where used
RDIMMs				
46W0784	A5B6	4GB TruDDR4 Memory (1Rx8, 1.2V) PC4-17000 CL15 2133MHz LP RDIMM	16	-
46W0788	A5B5	8GB TruDDR4 Memory (1Rx4, 1.2V) PC4-17000 CL15 2133MHz LP RDIMM	16	-
46W0792	A5B8	8GB TruDDR4 Memory (2Rx8, 1.2V) PC4-17000 CL15 2133MHz LP RDIMM	16	22x, 42x
46W0796	A5B7	16GB TruDDR4 Memory (2Rx4, 1.2V) PC4-17000 CL15 2133MHz LP RDIMM	16	62x
95Y4808	A5UJ	32GB TruDDR4 Memory (2Rx4, 1.2V) PC4-17000 CL15 2133MHz LP RDIMM	16	-
LRDIMMs				
46W0800	A5B9	32GB TruDDR4 Memory (4Rx4, 1.2V) PC417000 CL15 2133MHz LP LRDIMM	16	-
95Y4812	A5UK	64GB TruDDR4 Memory (4Rx4, 1.2V) PC4-17000 CL15 2133MHz LP LRDIMM		

In the nx360 M5, the maximum memory speed of a configuration is the lower of the following two values:

- ▶ Memory speed of the processor
- ▶ Memory speed of the DIMM

For optimal performance, use DIMMs in multiples of four use the four memory channels of the processor, and use the same number of DIMMs for each processor. That is, four or eight DIMMs when a single processor is installed, and eight or 16 DIMMs when two processors are installed.

The following memory protection technologies are supported:

- ▶ ECC
- ▶ Chipkill (x4 memory options only: 1Rx4, 2Rx4, and 4Rx4)
- ▶ Memory mirroring (planned for 3Q/2015)
- ▶ Memory rank sparing (planned for 3Q/2015)

If memory mirroring is used, DIMMs must be installed in pairs (minimum of one pair per CPU), and both DIMMs in a pair must be identical in type and size.

If memory rank sparing is used, a minimum of one quad-rank DIMM or two single-rank or dual-rank DIMMs must be installed per populated channel (the DIMMs do not need to be identical). In rank sparing mode, one rank of a DIMM in each populated channel is reserved as spare memory. The size of a rank varies depending on the DIMMs that are installed.

Table 4-7 on page 76 lists the maximum memory speeds that are achievable. The table also shows the maximum memory capacity at any speed that is supported by the DIMM and the maximum memory capacity at the rated DIMM speed.

Table 4-7 Maximum memory speeds

Spec	RDIMMs		LRDIMMs
Rank	Single rank	Dual rank	Quad rank
Part numbers	46W0784 (4 GB) 46W0788 (8 GB)	46W0792 (8 GB) 46W0796 (16 GB) 95Y4808 (32 GB)	46W0800 (32 GB) 95Y4812 (64 GB)
Rated speed	2133 MHz	2133 MHz	2133 MHz
Rated voltage	1.2 V	1.2 V	1.2 V
Operating voltage	1.2 V	1.2 V	1.2 V
Max quantity ^a	16	16	16
Largest DIMM	8 GB	16 GB	32 GB
Max memory capacity	128 GB	256 GB	512 GB
Max memory at rated speed	128 GB	256 GB	512 GB
Maximum operating speed (MHz)			
1 DIMM per channel	2133 MHz	2133 MHz	2133 MHz
2 DIMMs per channel	2133 MHz	2133 MHz	2133 MHz

a. The maximum quantity that is supported is shown for two installed processors. When one processor is installed, the maximum quantity that is supported is half of that shown.

DIMM installation order

The NeXtScale nx360 M5 boots with only one memory DIMM installed per processor. However, the suggested memory configuration is to balance the memory across all the memory channels on each processor to use the available memory bandwidth.

The locations of the DIMM sockets relative to the processors are shown in Figure 4-5.

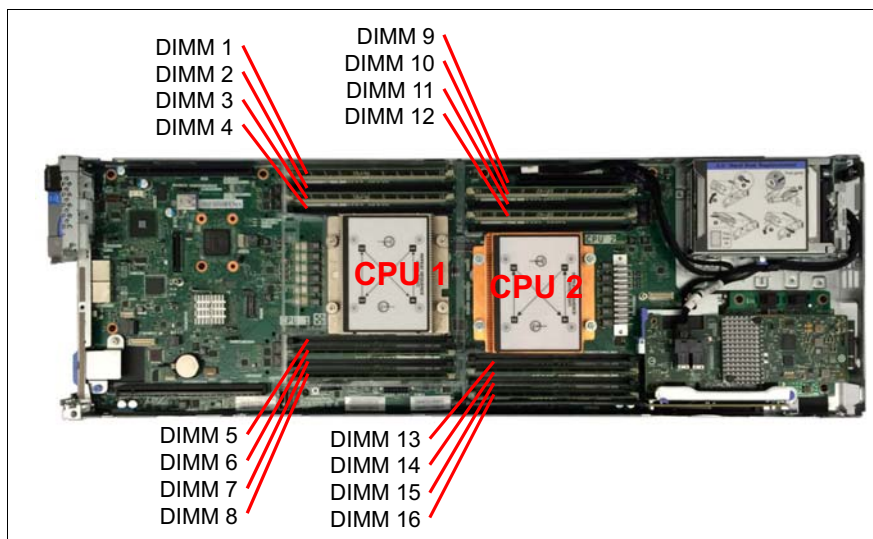


Figure 4-5 DIMM and processor numbering

Memory DIMM installation: Independent channel mode

Table 4-8 lists DIMM installation order if you have one installed processor.

Table 4-8 Memory population table with one processor installed (independent channel mode)

	Processor 1							
Number of DIMMs	DIMM 1	DIMM 2	DIMM 3	DIMM 4	DIMM 5	DIMM 6	DIMM 7	DIMM 8
1								x
2	x							x
3	x					x		x
4	x		x			x		x
5	x		x			x	x	x
6	x	x	x			x	x	x
7	x	x	x		x	x	x	x
8	x	x	x	x	x	x	x	x

Table 4-9 lists DIMM installation order if you have two installed processors. A minimum of two memory DIMMs (one for each processor) are required when two processors are installed.

Table 4-9 Memory population table with two processors installed (independent channel mode)

	Processor 1								Processor 2							
Number of DIMMs	DIMM 1	DIMM 2	DIMM 3	DIMM 4	DIMM 5	DIMM 6	DIMM 7	DIMM 8	DIMM 9	DIMM 10	DIMM 11	DIMM 12	DIMM 13	DIMM 14	DIMM 15	DIMM 16
2								x	x							
3	x							x	x							
4	x							x	x							x
5	x					x		x	x							x
6	x					x		x	x		x					x
7	x		x			x		x	x		x					x
8	x		x			x		x	x		x			x		x
9	x		x			x	x	x	x		x			x		x
10	x		x			x	x	x	x	x	x			x		x
11	x	x	x			x	x	x	x	x	x			x		x
12	x	x	x			x	x	x	x	x	x			x	x	x
13	x	x	x		x	x	x	x	x	x	x			x	x	x
14	x	x	x		x	x	x	x	x	x	x	x		x	x	x
15	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x
16	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Memory DIMM installation: Mirrored-channel mode

In mirrored channel mode, the channels are paired and both channels in a pair store the same data. Because of the redundancy, the effective memory capacity of the compute node is half the installed memory capacity.

The pair of DIMMs that are installed in each channel must be identical in capacity, type, and rank count.

Table 4-10 lists DIMM installation order if you have one installed processor.

Table 4-10 Memory population table with one processor installed (mirrored channel mode)

	Processor 1							
Number of DIMMs	DIMM 1	DIMM 2	DIMM 3	DIMM 4	DIMM 5	DIMM 6	DIMM 7	DIMM 8
2						x		x
4	x		x			x		x
6	x		x		x	x	x	x
8	x	x	x	x	x	x	x	x

Table 4-11 lists DIMM installation order if you have two installed processors. A minimum of four memory DIMMs (two for each processor) are required when two processors are installed.

Table 4-11 Memory population table with two processors installed (mirrored channel mode)

	Processor 1								Processor 2							
Number of DIMMs	DIMM 1	DIMM 2	DIMM 3	DIMM 4	DIMM 5	DIMM 6	DIMM 7	DIMM 8	DIMM 9	DIMM 10	DIMM 11	DIMM 12	DIMM 13	DIMM 14	DIMM 15	DIMM 16
4						x		x	x		x					
6	x		x			x		x	x		x					
8	x		x			x		x	x		x			x		x
10	x		x		x	x	x	x	x		x			x		x
12	x		x		x	x	x	x	x	x	x	x		x		x
14	x	x	x	x	x	x	x	x	x	x	x	x		x		x
16	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Memory DIMM installation: Rank-Sparing mode

In rank sparing mode, a minimum of one quad-rank DIMM or two single-rank or dual-rank DIMMs must be installed per populated channel (the DIMMs do not need to be identical). In rank sparing mode, one rank of a DIMM in each populated channel is reserved as spare memory. The size of a rank varies depending on the DIMMs that are installed.

Table 4-12 lists the single and dual-rank DIMM installation order if you have one installed processor.

Table 4-12 Memory population table with one processor installed (rank sparing mode)

	Processor 1							
Number of DIMMs	DIMM 1	DIMM 2	DIMM 3	DIMM 4	DIMM 5	DIMM 6	DIMM 7	DIMM 8
2							x	x
4	x	x					x	x
6	x	x			x	x	x	x
8	x	x	x	x	x	x	x	x

Table 4-13 lists the single and dual-rank DIMM installation order if you have two installed processors. A minimum of four single and dual-rank memory DIMMs (two for each processor) are required when two processors are installed.

Table 4-13 Memory population table with two processors installed (rank sparing mode)

	Processor 1								Processor 2							
Number of DIMMs	DIMM 1	DIMM 2	DIMM 3	DIMM 4	DIMM 5	DIMM 6	DIMM 7	DIMM 8	DIMM 9	DIMM 10	DIMM 11	DIMM 12	DIMM 13	DIMM 14	DIMM 15	DIMM 16
4							x	x	x	x						
6	x	x					x	x	x	x						
8	x	x					x	x	x	x					x	x
10	x	x			x	x	x	x	x	x					x	x
12	x	x			x	x	x	x	x	x	x	x			x	x
14	x	x	x	x	x	x	x	x	x	x	x	x			x	x
16	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

4.1.7 NeXtScale 12G Storage Native Expansion Tray

The NeXtScale 12G Storage Native Expansion Tray is a half-wide 1U expansion tray that attaches to the nx360 M5 to provide up to seven extra 3.5-inch simple-swap drives. The tray allows the configuration of storage-rich nx360 M5 compute nodes (up to 48 TB per node that uses 6 TB drives).

Note: The 12G Storage Native Expansion Tray and the PCIe Native Expansion Tray cannot be connected to the same compute node.

Figure 4-6 shows the NeXtScale 12G Storage Native Expansion Tray that is attached to a NeXtScale nx360 M5 node with the cover removed, which shows that seven 3.5-inch drives are installed.



Figure 4-6 NeXtScale 12G Storage Native Expansion Tray attached to an nx360 M5 compute node

Ordering information is listed in Table 4-14.

Table 4-14 Ordering information

Part number	Feature code	Description
00KG601	ASGR	NeXtScale 12G Storage Native Expansion Tray

Note: When the NeXtScale 12G Storage Native Expansion Tray is used, one of the following disk controller adapters must be installed in the front PCIe slot (slot 1) in the nx360 M5:

- ▶ ServeRAID M5210 SAS/SATA Controller for System x, 46C9110
- ▶ ServeRAID M1215 SAS/SATA Controller for System x, 46C9114
- ▶ N2215 SAS/SATA HBA for System x, 47C8675

No other PCIe adapter can be used for selection. The ML2 slot is still available.

4.1.8 Internal storage

The NeXtScale nx360 M5 server supports the following drives:

Drives internal to the server:

- ▶ One 3.5-inch simple-swap HDD
- ▶ Four 2.5-inch HDDs or SSDs (two simple-swap and two hot-swap)
- ▶ Four 1.8-inch simple-swap SSDs

In addition, with optional expansion trays:

- ▶ Seven additional 3.5-inch simple-swap HDDs with the use of the NeXtScale 12G Storage Native Expansion Tray (see 4.1.7, “NeXtScale 12G Storage Native Expansion Tray” on page 79), or

- Four additional 2.5-inch hot-swap HDDs or SSDs with the use of the PCIe 2U Native Expansion Tray

Simple-swap drives

The NeXtScale nx360 M5 server supports the following internal drives to be installed at the rear of the server:

- Up to one 3.5-inch simple-swap HDD
- Up to two 2.5-inch simple-swap HDDs or SSDs
- Up to four 1.8-inch simple-swap SSDs

Figure 4-7 shows the three variations.

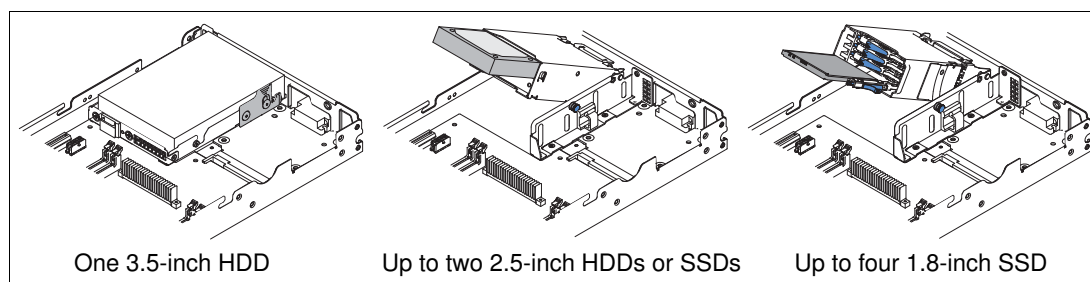


Figure 4-7 Drive cage options

These internal drives are installed in a drive cage. Ordering information for these drive cages are listed in Table 4-15.

Table 4-15 Internal drive cages for the drive bay in the nx360 M5

Part number	Feature code	Description	Models where used
00KA895	A5V3	nx360 M5 1.8-inch Rear Drive Cage	-
00KA894	A5V2	nx360 M5 2.5-inch Rear Drive Cage	62x
00FL465	A5K1	nx360 M5 3.5-inch Rear Drive Cage	22x, 42x
00KG603	ASGS	nx360 3.5-inch HDD8 Cage - HW RAID ^a	-

a. This single 3.5-inch drive cage is installed in the nx360 M5 when you have the 12G Storage Native Expansion Tray attached and want to configure all eight 3.5-inch drives as one RAID array.

In addition, the nx360 M5 supports seven more 3.5-inch drive bays if the NeXtScale 12G Storage Native Expansion Tray is attached. The 12G Storage Native Expansion Tray can be used with any of the rear drive cages listed in Table 4-15 to provide the following internal drive combinations:

- Up to eight 3.5-inch simple-swap SATA, NL SATA, or NL SAS drives
- Up to seven 3.5-inch simple-swap SATA, NL SATA, or NL SAS drives and two 2.5-inch simple-swap SATA drives
- Up to seven 3.5-inch simple-swap SATA, NL SATA, or NL SAS drives and four 1.8-inch simple-swap SATA SSDs

Drives that are used in the 12G Storage Native Expansion Tray do not need a cage.

There are two 3.5-inch drive cages (part numbers 00FL465 and 00KG603 in the last two rows of Table 4-15 on page 81). If the 12G Storage Native Expansion Tray is attached to the nx360 M5, the usage of the RAID cage (feature ASGS, option 00KG603) allows you to configure a RAID array that spans all eight drives; that is, the seven in the storage tray and the one drive internal to the nx360 M5. Such a configuration is connected to a RAID adapter or SAS HBA; the use of the internal SATA ports is not supported by this RAID cage.

If the 3.5-inch HDD cage (feature A5K1) is used, a RAID array can be formed only with the seven drives in the storage tray. In such a configuration, the drives in the storage tray are connected to RAID adapter or SAS HBA and the single drive in the nx360 M5 is connected to an onboard SATA port.

Hot-swap drives

If the internal drives are 2.5-inch drive bays (or if no internal drive bay is selected), the server also supports two more 2.5-inch drive bays. These bays are front accessible and are hot-swap drive bays. These hot-swap drive bays take the place of the full-height PCIe slot, as shown in Figure 4-8.

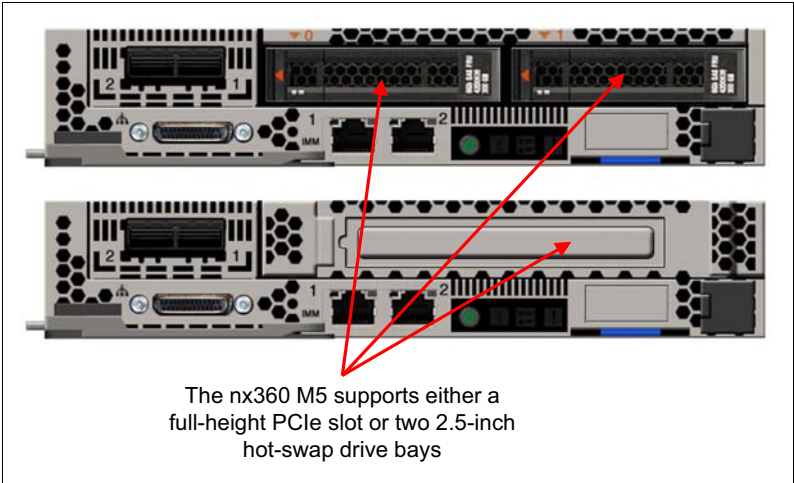


Figure 4-8 NeXtScale nx360 M5 configurations: Hot-swap 2.5-inch drive bays or full-height PCIe slot

Table 4-16 lists the ordering information for the two 2.5-inch hot-swap drive bays.

Table 4-16 Drive cage for the hot-swap drive bay in the nx360 M5

Part number	Feature code	Description	Models where used
00FL175	A5NA	nx360 M5 2.5" Front Hot Swap Drive Cage	-

Note: The NeXtScale 12G Storage native Expansion Tray cannot be used with the nx360 M5 2.5-inch Front Hot Swap Drive Cage.

Drive bays in the NeXtScale PCIe 2U Native Expansion Tray

As an alternative to the Storage Native Expansion Tray, the PCIe 2U Native Expansion Tray offers up to 4 additional 2.5-inch hot-swap drive bays beyond the bays internal to the server.

The following drive combinations are supported:

- ▶ One 3.5-inch simple-swap SATA, NL SATA or NL SAS drive (internal) and up to four 2.5-inch hot-swap SAS/SATA drives (expansion tray)
- ▶ Two 2.5-inch simple-swap SATA drives (internal) and up to four 2.5-inch hot-swap SAS/SATA drives (expansion tray)
- ▶ Four 1.8-inch simple-swap SATA SSDs (internal) and up to four 2.5-inch hot-swap SAS/SATA drives (expansion tray)

The use of the four 2.5-inch hot-swap drive bays in the expansion tray require an optional backplane as listed in Table 4-17.

Table 4-17 Table 32. Backplane for the hot-swap drive bays in the PCIe 2U Native Expansion Tray

Part number	Feature code	Description	Models where used
44X4104	A4A6	4x 2.5" HDD Riser (Backplane and SAS cable)	-

These drive bays in the require a RAID controller or SAS HBA installed in the dedicated RAID slot at rear of the server.

For more information see 4.1.15, “NeXtScale PCIe 2U Native Expansion Tray” on page 98

4.1.9 Controllers for internal storage

The onboard SATA controller (integrated into the Intel C612 chipset) supports any of the following drive configurations:

- ▶ One 3.5-inch simple-swap SATA or NL SATA drive
- ▶ Up to two 2.5-inch simple-swap NL SATA drives
- ▶ Up to four 1.8-inch SATA Enterprise Value SSDs

The following drive combinations can be used instead with a RAID controller or SAS/SATA HBA that is installed in the internal RAID adapter riser slot:

- ▶ Up to two 2.5-inch simple-swap NL SATA drives
- ▶ Up to four 1.8-inch SATA Enterprise Value SSDs

Any of the following drive configurations require a RAID controller or SAS/SATA HBA that is installed in the internal RAID adapter riser slot:

- ▶ Up to two 2.5-inch simple-swap SAS drives
- ▶ Up to two 2.5-inch hot-swap drives (installed in the front drive bays)

The supported RAID controller or SAS/SATA host bus adapter is installed in a dedicated RAID adapter slot (through a riser card) at the rear of the server that is next to the internal drive bays. Installation of the adapter is shown in Figure 4-9.

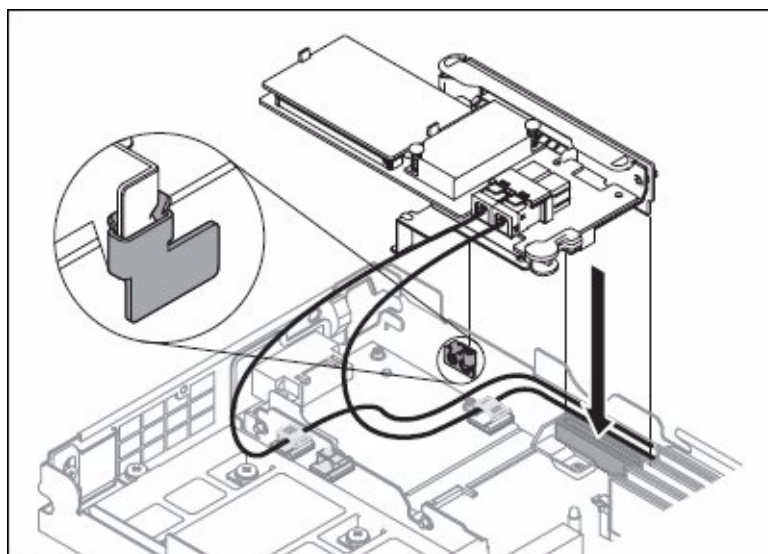


Figure 4-9 RAID controller and RAID riser card installation

When the NeXtScale 12G Storage Native Expansion Tray is used, the RAID controller or SAS/SATA host bus adapter must be installed in the front PCIe slot (slot 1) in the nx360 M5. In this case, the drives in the storage tray are connected to the RAID controller or SAS/SATA host bus adapter, the 1.8-inch and 2.5-inch simple-swap rear drives of the nx360 M5 are connected to the onboard SATA controller and the 3.5-inch simple-swap rear drive of the nx360 M5 can be connected to the on board SATA controller or to the RAID controller or SAS/SATA host bus adapter. For more information about connectivity rules, see Table 4-20 on page 86.

Table 4-18 lists the RAID controllers and HBAs that are used for internal disk storage of the nx360 M5 server and the Riser card that is needed to install the adapter.

Table 4-18 Drive controllers for internal storage

Part number	Feature code	Description	Maximum supported
Riser card for RAID adapter			
00FL179	A5JZ	nx360 M5 RAID Riser	1
RAID Controllers and SAS Host Bus Adapters			
46C9110	A3YZ	ServeRAID M5210 SAS/SATA Controller for System x	1 ^a
46C9114	A45W	ServeRAID M1215 SAS/SATA Controller for System x	1 ^a
47C8675	A3YY	N2215 SAS/SATA HBA for System x	1 ^a
Hardware Upgrades: ServeRAID M5210			
47C8656	A3Z0	ServeRAID M5200 Series 1GB Cache/RAID 5 Upgrade	1
47C8660	A3Z1	ServeRAID M5200 Series 1GB Flash/RAID 5 Upgrade	1

Part number	Feature code	Description	Maximum supported
47C8664	A3Z2	ServeRAID M5200 Series 2GB Flash/RAID 5 Upgrade	1
47C8668	A3Z3	ServeRAID M5200 Series 4GB Flash/RAID 5 Upgrade	1
Features on Demand Upgrades: ServeRAID M5210			
47C8706	A3Z5	ServeRAID M5200 Series RAID 6 Upgrade-FoD	1
47C8708	A3Z6	ServeRAID M5200 Series Zero Cache/RAID 5 Upgrade-FoD	1
47C8710	A3Z7	ServeRAID M5200 Series Performance Accelerator-FoD	1
47C8712	A3Z8	ServeRAID M5200 Series SSD Caching Enabler-FoD	1
Features on Demand Upgrades: ServeRAID M1215			
00AE930	A5H5	ServeRAID M1200 Zero Cache/RAID 5 Upgrade FOD	1

a. Mutually exclusive. Only one of these adapters is supported in the nx360 M5 and it requires RAID Riser 00FL179 when it is installed in the internal RAID slot.

Controller specifications

The ServeRAID M5210 SAS/SATA Controller includes the following specifications:

- ▶ Eight internal 12 Gbps SAS/SATA ports
- ▶ 12 Gbps throughput per port
- ▶ Based on the LSI SAS3108 12 Gbps ROC controller
- ▶ Two mini-SAS HD internal connectors (SFF8643)
- ▶ Supports connections to SAS/SATA drives and SAS Expanders
- ▶ Supports RAID levels 0, 1, and 10
- ▶ Supports RAID levels 5 and 50 with optional M5200 Series RAID 5 upgrades
- ▶ Supports RAID 6 and 60 with the optional M5200 Series RAID 6 Upgrade
- ▶ Supports 1 GB cache (no battery backup) or 1 GB or 2 GB flash-backed cache
- ▶ Supports performance upgrades through Features on Demand

The ServeRAID M1215 SAS/SATA Controller has the following specifications:

- ▶ Eight internal 12 Gbps SAS/SATA ports
- ▶ Up to 12 Gbps throughput per port
- ▶ Two internal mini-SAS HD connectors (SFF8643)
- ▶ Based on the LSI SAS3008 12 Gbps RAID on Chip (ROC) controller
- ▶ Support for RAID levels 0, 1, and 10 standard; support for RAID 5 and 50 with optional FoD upgrade
- ▶ Zero Controller Cache, no battery/flash backup
- ▶ Optional support for self-encrypting drives (SEDs) with MegaRAID SafeStore (with RAID 5 upgrade)
- ▶ Fixed stripe size of 64 KB

For more information about these RAID controllers, see the list of Product Guides in the RAID adapters category at this website:

<http://lenovopress.com/systemx/raid>

The N2215 SAS/SATA HBA features the following specifications:

- ▶ Eight internal 12 Gbps SAS/SATA ports (support for 12, 6, or 3 Gbps SAS speeds and 6 or 3 Gbps SATA speeds)
- ▶ Up to 12 Gbps throughput per port
- ▶ Two internal x4 HD Mini-SAS connectors (SFF-8643)
- ▶ Based on the LSI SAS3008 12 Gbps controller
- ▶ Non-RAID (JBOD mode) support for SAS and SATA HDDs and SSDs (RAID not supported)
- ▶ PCI low profile, half-length MD2 form factor
- ▶ PCI Express 3.0 x8 host interface
- ▶ Optimized for SSD performance
- ▶ High-performance IOPS LSI Fusion-MPT architecture
- ▶ Advanced power management support
- ▶ Support for SSP, SMP, STP, and SATA protocols
- ▶ End-to-End CRC with Advanced Error Reporting
- ▶ T-10 Protection Model for early detection of and recovery from data corruption
- ▶ Spread Spectrum Clocking for EMI reductions

For more information about this SAS/SATA HBA, see the list of Product Guides in the host bus adapters category at this website:

<http://lenovopress.com/systemx/hba>

Table 4-19 lists the SAS/SATA cables that are supported for connection from the onboard SATA controller, RAID controller, or SAS/SATA HBA to the various drive cages.

Table 4-19 SAS/SATA Cables

Part number	Feature code	Description
00FL170	A5K3	nx360 M5 1x2, 2.5-inch 12G HDD short cable, HW RAID (stack-up)
00FL173	A5K7	nx360 M5 1.8-inch SSD 12G short cable vertical (HW RAID)
00KA360	A5QH	nx360 M5 1x2, 2.5-inch 12G HDD short cable, HW RAID (stack-up) Port 1
00FL466	A5K4	nx360 M5 2.5-inch HDD 2x cable right angle cable (no RAID)
00FL467	A5K5	nx360 M5 Rear SSD cable 1.8-inch server node 4 SSD to planar (no RAID)

Table 4-20 lists the required combination of drive cage, SAS/SATA adapter, and SAS/SATA cable that is based on the selected drive type.

Table 4-20 Required drive cages and cables based on drive type and adapter

Drive type	Max drive qty	Cage feature code	Cable feature code No RAID		Cable feature code Hardware RAID	
			On board SATA	N2215 HBA	ServeRAID M1215	ServeRAID M5210
Simple-Swap rear drives only						
1.8-inch SS SATA SSD	4	A5V3	A5K5	A5K7	A5K7	A5K7

Drive type	Max drive qty	Cage feature code	Cable feature code No RAID		Cable feature code Hardware RAID	
			On board SATA	N2215 HBA	ServeRAID M1215	ServeRAID M5210
2.5-inch SS SATA HDD	2	A5V2	A5K4	A5K3	A5K3	A5K3
2.5-inch SS SATA SSD	2	A5V2	A5K4	A5K3	A5K3	A5K3
2.5-inch SS SAS HDD	2	A5V2	No support	A5K3	A5K3	A5K3
2.5-inch SS SAS SED HDD	2	A5V2	No support	No support	A5K3	A5K3
3.5-inch SS SATA HDD	1	A5K1	With cage ^a	No support	No support	No support
3.5-inch SS NL SATA HDD	1	A5K1	With cage ^a	No support	No support	No support
3.5-inch SS NL SAS HDD	0	No support	No support	No support	No support	No support
3.5-inch SS NL SAS SED HDD	0	No support	No support	No support	No support	No support
Host-Swap drives only						
2.5-inch HS SAS HDD	2	A5NA	No support	With cage ^a	With cage ^a	With cage ^a
2.5-inch HS NL SATA HDD	2	A5NA	No support	With cage ^a	With cage ^a	With cage ^a
2.5-inch HS SATA SSD	2	A5NA	No support	With cage ^a	With cage ^a	With cage ^a
2.5-inch HS SAS SSD	2	A5NA	No support	With cage ^a	With cage ^a	With cage ^a
2.5-inch HS SAS SED HDD	2	A5NA	No support	No support	With cage ^a	With cage ^a
2.5-inch HS SAS SED SSD	2	A5NA	No support	No support	With cage ^a	With cage ^a
Mixed Simple-Swap and Hot-Swap drives^b						
2.5-inch Drives	2 + 2	A5V2 + A5NA	No support	A5QH ^c	A5QH ^c	A5QH ^c
2.5-inch Drives SED ^d	2 + 2	A5V2 + A5NA	No support	No support	A5QH ^c	A5QH ^c
With 12G Storage Native Expansion Tray attached (adds 7x 3.5-inch bays) (two controllers)^e						
1.8-inch SS SSD	4 + 7	A5V3	A5K5 ^f	With tray ^g	With tray ^g	With tray ^g
2.5-inch SS (NL) SATA HDD	2 + 7	A5V2	A5K4 ^f	With tray ^g	With tray ^g	With tray ^g
2.5-inch SS (NL) SATA SSD	2 + 7	A5V2	A5K4 ^f	With tray ^g	With tray ^g	With tray ^g
3.5-inch SS (NL) SATA HDD	1 + 7	A5K1	With cage ^a	With tray ^g	With tray ^g	With tray ^g
3.5-inch SS NL SAS HDD	0 + 7	None	None	With tray ^g	With tray ^g	With tray ^g
3.5-inch SS NL SAS SED	0 + 7	None	None	No support	With tray ^g	With tray ^g
With 12G Storage Native Expansion Tray attached (adds 7x 3.5-inch bays) (one controller for all drives)						
3.5-inch Drive	8	ASGS + ASGR	No support	With cage ^a & tray ^g	With cage ^a & tray ^g	With cage ^a & tray ^g
3.5-inch Drive SED	8	ASGS + ASGR	No support	No support	With cage ^a & tray ^g	With cage ^a & tray ^g

a. Cable is provided with the cage.

b. Any combination of 2.5-inch drive is allowed if no hardware RAID is required. For all four disks to be part of the same RAID array, they must have similar interface, capacity, and speed. Alternatively, two pairs of similar drives can be used to be part of two separate RAID arrays.

- c. The second cable is provided with the hot-swap cage.
- d. At least one of the drives is an SED.
- e. The two controllers must be onboard SATA for internal rear SS drives, and N2215 or ServeRAID M1215 or ServeRAID M5210 for drives in the storage tray.
- f. Cable that is used to connect the rear SS drives to the onboard SATA controller.
- g. Cable is provided with the tray to connect to the controller.

4.1.10 Internal drive options

Table 4-21 on page 89 through to Table 4-26 on page 91 list drive options for the internal disk storage of the nx360 M5 server. Consider the following rules for mixing drive types:

- The server supports the following drive types:
 - SATA, NL SATA, SAS and NL SAS HDDs
 - SAS and SATA SSDs
 - SAS and NL SAS SED HDDs
 - SAS SED SSDs

These drive types can be intermixed in a server and on the same RAID controller, but they cannot be intermixed in the same RAID array. That is, all drives in a single RAID array must be all SAS (and NL SAS) or all SATA (and NL SATA), and must have the same size and speed.

- Mixing hot-swap (front) and simple-swap (internal) drives: Only 2.5-inch simple-swap drives are supported in combination with 2.5-inch hot-swap drives. The 1.8-inch and 3.5-inch internal drives are not supported when 2.5-inch hot-swap drives are installed.

Tip: To mix hot-swap (front) and simple-swap (internal) HDDs with x-config configurator tool, you must click **Split** in the SFF Slim SAS SATA section.

Self-encrypting drives

Table 4-21 on page 89 through to Table 4-26 on page 91 list a number of SEDs. To use these drives, they must be combined with a compatible RAID controller.

The SED are selectable with ServeRAID M5210 or ServeRAID M1215 controllers. Also, either of the following RAID Upgrade or FOD is required:

- ServeRAID M5210 SAS/SATA Controller for System x (PN 46C9110 / FC A3YZ) upgrades:
 - ServeRAID M5210 1GB Cache RAID 5 Upgrade (PN 47C8656 / FC A3Z0);
 - ServeRAID M5210 1GB Flash RAID 5 Upgrade (PN 47C8660 / FC A3Z1);
 - ServeRAID M5210 2GB Flash RAID 5 Upgrade (PN 47C8664 / FC A3Z2);
 - ServeRAID M5210 4GB Flash RAID 5 Upgrade (PN 47C8668 / FC A3Z3); or
 - ServeRAID M5200 Series Zero Cache/RAID 5 Upgrade (PN 47C8708 / FC A3Z6)
- ServeRAID M1215 SAS/SATA Controller for System x (PN 46C9114 / FC A45W) upgrade:
 - ServeRAID M1200 Zero Cache/RAID 5 Upgrade FOD (PN 00AE930 / FC A5H5)

In addition, FC 5977 (no Lenovo configured RAID required) must be selected. The RAID configuration is not performed in the manufacturing plant with self-encrypting drives.

Simple swap drives

The supported simple-swap drives are listed in the following tables:

- Table 4-21 on page 89: 3.5-inch simple-swap HDDs
- Table 4-22 on page 89: 2.5-inch simple-swap HDDs

- Table 4-23 on page 90: 2.5-inch simple-swap SSDs
- Table 4-24 on page 90: 1.8-inch simple-swap SSDs

Table 4-21 Disk drive options for internal disk storage (3.5-inch simple-swap HDDs)

Part number	Feature code	Description	Maximum supported ^a
3.5-inch 6 Gb SATA simple-swap HDDs			
00AD010	A487	1TB 7.2K 6Gbps SATA 3.5-inch HDD for NeXtScale System	1 / 8
00AD025	A4GC	4TB 7.2K 6Gbps SATA 3.5-inch HDD for NeXtScale System	1 / 8
3.5-inch 6 Gb NL SATA simple-swap HDDs: 512e format			
00FN123	A5VV	2TB 7.2K 6Gbps NL SATA 3.5-inch 512e HDD for NeXtScale System	1 / 8
00FN168	A5VY	5TB 7.2K 6Gbps NL SATA 3.5-inch 512e HDD for NeXtScale System	1 / 8
00FN183	A5VZ	6TB 7.2K 6Gbps NL SATA 3.5-inch 512e HDD for NeXtScale System	1 / 8
3.5-inch 6 Gb NL SAS simple-swap HDDs: 512e format			
00FN213	AS4J	4TB 7.2K 12Gbps NL SAS 3.5-inch 512e HDD for NeXtScale System	1 / 8
3.5-inch 6 Gb NL SAS simple-swap SEDs: 512e format			
00FN243	AS4E	2TB 7.2K 12Gbps NL SAS 3.5-inch 512e SED for NeXtScale System	1 / 8
00FN263	AS4G	6TB 7.2K 12Gbps NL SAS 3.5-inch 512e SED for NeXtScale System	1 / 8

a. One drive supported if the 12G Storage Native Expansion Tray is not attached; 8 drives supported if the 12G Storage Native Expansion Tray is attached.

Table 4-22 Disk drive options for internal disk storage (2.5-inch simple-swap HDDs)

Part number	Feature code	Description	Maximum supported
2.5-inch 12Gb SAS 15K simple-swap HDDs: 512e format			
00NA331	ASBZ	300GB 15K 12Gbps SAS 2.5-inch 512e HDD for NeXtScale System	2
00NA336	ASC0	600GB 15K 12Gbps SAS 2.5-inch 512e HDD for NeXtScale System	2
2.5-inch 12Gb SAS 10K simple-swap HDDs: 512e format			
00NA341	ASC1	600GB 10K 12Gbps SAS 2.5-inch 512e HDD for NeXtScale System	2
00NA346	ASC2	900GB 10K 12Gbps SAS 2.5-inch 512e HDD for NeXtScale System	2
00NA351	ASC3	1.2TB 10K 12Gbps SAS 2.5-inch 512e HDD for NeXtScale System	2
00NA356	ASC4	1.8TB 10K 12Gbps SAS 2.5-inch 512e HDD for NeXtScale System	2
2.5-inch 12Gb SAS 10K simple-swap SEDs: 512e format			
00NA376	ASC8	900GB 10K 12Gbps SAS 2.5-inch 512e SED for NeXtScale System	2
2.5-inch 6 Gb SAS simple-swap HDDs			
00AD055	A48D	300GB 10K 6Gbps SAS 2.5-inch HDD for NeXtScale System	2
00AD060	A48F	600GB 10K 6Gbps SAS 2.5" HDD for NeXtScale System	2
00AJ290	A5NG	600GB 15K 6Gbps SAS 2.5-inch' HDD for NeXtScale System	2

Part number	Feature code	Description	Maximum supported
00FN040	A5NC	1.2TB 10K 6Gbps SAS 2.5-inch HDD for NeXtScale System	2

Table 4-23 Disk drive options for internal disk storage (2.5-inch simple-swap SSDs)

Part number	Feature code	Description	Maximum supported
00FN020	A57K	120GB SATA 2.5-inch MLC Enterprise Value SSD for NeXtScale System	2
00FN025	A57L	240GB SATA 2.5-inch MLC Enterprise Value SSD for NeXtScale System	2
00FN030	A57M	480GB SATA 2.5-inch MLC Enterprise Value SSD for NeXtScale System	2
00FN035	A57N	800GB SATA 2.5-inch MLC Enterprise Value SSD for NeXtScale System	2
00FN293	A5U8	S3500 1.6TB SATA 2.5-inch MLC Enterprise Value SSD for NeXtScale System	2

Table 4-24 Disk drive options for internal disk storage (1.8-inch simple-swap SSDs)

Part number	Feature code	Description	Maximum supported
1.8-inch simple-swap Enterprise Value SSDs			
00AJ040	A4KV	S3500 80GB SATA 1.8-inch MLC Enterprise Value SSD for System x	4
00AJ050	A4KX	S3500 400GB SATA 1.8-inch MLC Enterprise Value SSD for System x	4

Hot-swap drives

The nx360 M5 optionally supports two hot-swap drive bays at the front of the server with the addition of the 2.5-inch Front Hot Swap Drive Cage (part number 00FL175) as listed in Table 4-16 on page 82. The supported hot-swap drives are listed in the following tables:

- ▶ Table 4-25: 2.5-inch hot-swap HDDs
- ▶ Table 4-26 on page 91: 2.5-inch hot-swap SSDs

Table 4-25 Disk drive options for internal disk storage (2.5-inch hot-swap HDDs)

Part number	Feature code	Description	Maximum supported ^a
2.5-inch 12Gb SAS 15K hot-swap HDDs: 512e format			
00NA221	ASBB	300GB 15K 12Gbps SAS 2.5-inch G3HS 512e HDD	2 / 4
00NA231	ASBD	600GB 15K 12Gbps SAS 2.5-inch G3HS 512e HDD	2 / 4
2.5-inch 12Gb SAS 10K hot-swap HDDs: 512e format			
00NA241	ASBF	600GB 10K 12Gbps SAS 2.5-inch G3HS 512e HDD	2 / 4
00NA251	ASBH	900GB 10K 12Gbps SAS 2.5-inch G3HS 512e HDD	2 / 4
00NA261	ASBK	1.2TB 10K 12Gbps SAS 2.5-inch G3HS 512e HDD	2 / 4
00NA271	ASBM	1.8TB 10K 12Gbps SAS 2.5-inch G3HS 512e HDD	2 / 4
2.5-inch 12Gb SAS 15K hot-swap SEDs: 512e format			
00NA281	ASBP	300GB 15K 12Gbps SAS 2.5-inch G3HS 512e SED	2 / 4

Part number	Feature code	Description	Maximum supported ^a
00NA286	ASBQ	600GB 15K 12Gbps SAS 2.5-inch G3HS 512e SED	2 / 4
2.5-inch 12Gb SAS 10K hot-swap SEDs: 512e format			
00NA291	ASBR	600GB 10K 12Gbps SAS 2.5-inch G3HS 512e SED	2 / 4
00NA301	ASBT	1.2TB 10K 12Gbps SAS 2.5-inch G3HS 512e SED	2 / 4
00NA306	ASBU	1.8TB 10K 12Gbps SAS 2.5-inch G3HS 512e SED	2 / 4
2.5-inch 6Gb SAS hot-swap HDDs			
00AJ096	A4TL	300GB 10K 6Gbps SAS 2.5-inch G3HS HDD	2 / 4
00AJ091	A4TM	600GB 10K 6Gbps SAS 2.5" G3HS HDD	2 / 4
00AJ126	A4TS	600GB 15K 6Gbps SAS 2.5-inch G3HS HDD	2 / 4
00AJ146	A4TP	1.2TB 10K 6Gbps SAS 2.5-inch G3HS HDD	2 / 4
2.5-inch 6Gb NL SATA hot-swap HDDs			
00AJ136	A4TW	500GB 7.2K 6Gbps NL SATA 2.5-inch G3HS HDD	2 / 4
00AJ141	A4TX	1TB 7.2K 6Gbps NL SATA 2.5" G3HS HDD	2 / 4

a. Maximum quantity is 2 if installed in the front-accessible bays inside the server, or 4 if installed in the NeXtScale PCIe 2U Native Expansion Tray

Table 4-26 Disk drive options for internal disk storage (2.5-inch hot-swap SSDs)

Part number	Feature code	Description	Maximum supported ^a
2.5-inch 12Gb SAS hot-swap Enterprise Capacity SSDs			
00NA671	ASW6	3.84TB 6Gb SAS Enterprise Capacity G3HS MLC SSD	2 / 4
2.5-inch 12Gb SAS hot-swap Enterprise SSDs			
00FN379	AS7C	200GB 12G SAS 2.5-inch MLC G3HS Enterprise SSD	2 / 4
00FN389	AS7E	400GB 12G SAS 2.5-inch MLC G3HS Enterprise SSD	2 / 4
00FN399	AS7G	800GB 12G SAS 2.5-inch MLC G3HS Enterprise SSD	2 / 4
00FN409	AS7J	1.6TB 12G SAS 2.5-inch MLC G3HS Enterprise SSD	2 / 4
2.5-inch 12Gb SAS hot-swap Enterprise SED SSDs			
00FN419	AS7L	400GB SED 12G SAS 2.5-inch MLC G3HS Enterprise SSD	2 / 4
00FN424	AS7M	800GB SED 12G SAS 2.5-inch MLC G3HS Enterprise SSD	2 / 4
2.5-inch 6Gb hot-swap Enterprise Value SSDs			
00AJ395	A577	120GB SATA 2.5-inch MLC G3HS Enterprise Value SSD	2 / 4
00AJ400	A578	240GB SATA 2.5-inch MLC G3HS Enterprise Value SSD	2 / 4
00AJ405	A579	480GB SATA 2.5-inch MLC G3HS Enterprise Value SSD	2 / 4
00AJ410	A57A	800GB SATA 2.5-inch MLC G3HS Enterprise Value SSD	2 / 4

Part number	Feature code	Description	Maximum supported ^a
00FN278	A5U6	S3500 1.6TB SATA 2.5-inch MLC G3HS Enterprise Value SSD	2 / 4

a. Maximum quantity is 2 if installed in the front-accessible bays inside the server, or 4 if installed in the NeXtScale PCIe 2U Native Expansion Tray

4.1.11 I/O expansion options

The NeXtScale nx360 M5 offers the following I/O expansion options:

- ▶ One PCIe 3.0 x16 ML2 adapter slot (optional, front accessible)
- ▶ One PCIe 3.0 x16 full-height half-length slot (optional, front accessible)
- ▶ One PCIe 2.0 x8 slot for internal RAID controller (optional, not front accessible)

Consider the following points:

- ▶ Each slot requires a riser card, as listed in Table 4-27.
- ▶ The use of the PCIe full-height slot and the use of the two 2.5-inch hot-swap drive bays are mutually exclusive.
- ▶ When the PCIe Native Expansion Tray is installed, the full-height half-length slot becomes a PCIe 3.0 x8 interface (physically still a x16 connector).
- ▶ When the 12G Storage Native Expansion Tray is installed, the front accessible PCIe slot is used to host the RAID controller or the SAS/SATA HBA.

The front accessible slots are shown in Figure 4-10. The internal slot for the RAID controller is shown in Figure 4-3 on page 66.

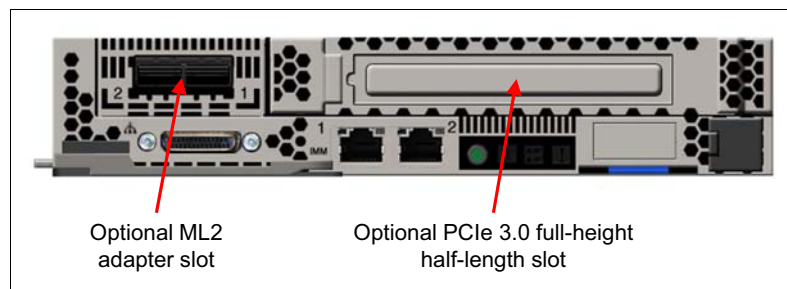


Figure 4-10 Optional front accessible PCIe slots

With the addition of the NeXtScale PCIe Native Expansion Tray, the server has two more PCIe 3.0 x16 full-height full-length double-width slots, as described in the 4.1.14, “NeXtScale PCIe Native Expansion Tray” on page 96.

The ordering information for optional risers for the three slots is listed in Table 4-27.

Table 4-27 Riser card options

Part number	Feature code	Description	Maximum supported
00FL180	A5JV	nx360 M5 ML2 Riser	1
00FL464	A5JY	nx360 M5 Compute Node Front Riser	1
00FL179	A5JZ	nx360 M5 RAID Riser	1

4.1.12 Network adapters

The NeXtScale nx360 M5 offers two Gigabit Ethernet ports with the following features:

- ▶ Broadcom BCM5717 Gigabit Ethernet controller
- ▶ TCP/IP Offload Engine (TOE) support
- ▶ Wake on LAN support
- ▶ Receive side Scaling (RSS) and Transmit side Scaling (TSS) support
- ▶ MSI and MSI-X capability (up to five MSI-X vectors)
- ▶ VLAN tag support (IEEE 802.1Q)
- ▶ Layer 2 priority encoding (IEEE 802.1p)
- ▶ Link aggregation (IEEE 802.3ad)
- ▶ Full-duplex flow control (IEEE 802.3x)
- ▶ IP, TCP, and UDP checksum offload (hardware based) on Tx/Rx over IPv4/IPv6
- ▶ Hardware TCP segmentation offload over IPv4/IPv6
- ▶ Jumbo frame support
- ▶ NIC Teaming (Load Balancing and Failover)
- ▶ One port that is shared with IMM2 by using the Network Controller-Sideband Interface (NC-SI)

The nx360 M5 server supports a Mezzanine LOM Generation 2 (ML2) adapter with a dedicated slot at the front of the server, as shown in Figure 4-10 on page 92. The use of an ML2 adapter also requires the installation of the ML2 Riser card. The Riser card and supported adapters are listed in Table 4-28.

Table 4-28 Mezzanine LOM Gen 2 (ML2) Adapters

Part number	Feature code	Description
Riser card for ML2 adapters		
00FL180	A5JV	nx360 M5 ML2 Riser
ML2 Ethernet adapters		
00D2026	A40S	Broadcom NetXtreme II ML2 Dual Port 10GbaseT for System x
00D2028	A40T	Broadcom NetXtreme II ML2 Dual Port 10GbE SFP+ for System x
00D1996	A40Q	Emulex VFA5 ML2 Dual Port 10GbE SFP+ Adapter for System x
00FP650	A5RK	Mellanox ConnectX-3 Pro ML2 2x40GbE/FDR VPI Adapter for System x
ML2 InfiniBand adapters		
00FP650	A5RK	Mellanox ConnectX-3 Pro ML2 2x40GbE/FDR VPI Adapter for System x
FCoE / iSCSI upgrades - Features on Demand		
00D8544	A4NZ	Emulex VFA5 ML2 FCoE/iSCSI License for System x (FoD) Features on Demand upgrade for 00D1996

Table 4-29 lists other supported network adapters in the standard full-height half-length PCIe slot. The use of an adapter in this slot also requires the installation of the PCIe Riser card.

Table 4-29 Network adapters

Part number	Feature code	Description
Riser card for PCIe adapters		
00FL464	A5JY	nx360 M5 Compute Node Front Riser
40 Gb Ethernet		
00D9550	A3PN	Mellanox ConnectX-3 40GbE / FDR IB VPI Adapter for System x
10 Gb Ethernet		
44T1370	A5GZ	Broadcom NetXtreme 2x10GbE BaseT Adapter for System x
00JY830	A5UU	Emulex VFA5 2x10 GbE SFP+ Adapter and FCoE/iSCSI SW for System x
00JY820	A5UT	Emulex VFA5 2x10 GbE SFP+ PCIe Adapter for System x
00JY824	A5UV	Emulex VFA5 FCoE/iSCSI SW for PCIe Adapter for System x (FoD)
49Y7960	A2EC	Intel X520 Dual Port 10GbE SFP+ Adapter
49Y7970	A2ED	Intel X540-T2 Dual Port 10GBaseT Adapter
81Y3520	AS73	Intel X710 2x10GbE SFP+ Adapter for System x
00D9690	A3PM	Mellanox ConnectX-3 10 GbE Adapter for System x
00FP650	A5RK	Mellanox ConnectX-3 Pro ML2 2x40GbE/FDR VPI Adapter for System x
47C9952	A47H	Solarflare SFN5162F 2x10GbE SFP+ Performant Adapter
Gigabit Ethernet		
94Y5180	A4Z6	Broadcom NetXtreme Dual Port 10GbE SFP+ Adapter for System x
90Y9370	A2V4	Broadcom NetXtreme I Dual Port GbE Adapter for System x
90Y9352	A2V3	Broadcom NetXtreme I Quad Port GbE Adapter for System x
00AG500	A56K	Intel I350-F1 1xGbE Fiber Adapter
00AG510	A56L	Intel I350-T2 2xGbE BaseT Adapter
00AG520	A56M	Intel I350-T4 4xGbE BaseT Adapter
InfiniBand		
00D9550	A3PN	Mellanox ConnectX-3 40GbE / FDR IB VPI Adapter for System x

More network adapters are offered as part of the Intelligent Cluster program, as listed in Table 4-30.

Table 4-30 Intelligent Cluster network adapters

Part number	Feature code	Description
100 Gb Ethernet		
None	ASWQ	Mellanox ConnectX-4 EDR IB VPI Single-port x16 PCIe 3.0 HCA
None	ASWR	Mellanox ConnectX-4 EDR IB VPI Dual-port x16 PCIe 3.0 HCA

Part number	Feature code	Description
40 Gb Ethernet		
46W0620	A4H5	Chelsio T580-LP-CR Dual-port (QSFP+) 40GbE PCI-E 3.0 Adapter
95Y3459	A2F8	Mellanox ConnectX-3 EN Dual-port QSFP+ 40GbE Adapter
00W0037	A2YE	Mellanox ConnectX-3 VPI Single-port QSFP FDR IB HCA
00W0041	A2YF	Mellanox ConnectX-3 VPI Dual-port QSFP FDR IB HCA
10 Gb Ethernet		
46W0609	A4H3	Chelsio T520-LL-CR Dual-port (SFP+) 10GbE PCI-E 3.0 Adapter
46W0615	A4H4	Chelsio T540-CR Quad-port (SFP+) 10GbE PCI-E 3.0 Adapter
00AE047	A4K1	Mellanox ConnectX-3 EN Single-port SFP+ 10GbE Adapter
00W0053	A2ZQ	Mellanox ConnectX-3 EN Dual-port SFP+ 10GbE Adapter
InfiniBand		
59Y1888	5763	Intel QLE7340 single-port 4X QDR IB x8 PCI-E 2.0 HCA
00D1773	AS97	Mellanox Connect-IB FDR IB Single-port PCIe 3.0 x16 HCA
00D1864	AS98	Mellanox Connect-IB FDR IB Single-port PCIe 3.0 x8 HCA
46W0571	A44E	Mellanox Connect-IB Dual-port QSFP FDR IB PCI-E 3.0 x16 HCA
00W0037	A2YE	Mellanox ConnectX-3 VPI Single-port QSFP FDR IB HCA
00W0041	A2YF	Mellanox ConnectX-3 VPI Dual-port QSFP FDR IB HCA
None	ASWQ	Mellanox ConnectX-4 EDR IB VPI Single-port x16 PCIe 3.0 HCA
None	ASWR	Mellanox ConnectX-4 EDR IB VPI Dual-port x16 PCIe 3.0 HCA
PCIe expansion		
None	ASYS	One Stop Switched x16 PCIe 3.0 HIB

For more information, see the list of Product Guides in the Network adapters category at this website:

<http://lenovopress.com/systemx/networkadapters>

4.1.13 Storage host bus adapters

Table 4-31 lists the storage HBAs that are supported by the nx360 M5 server. These HBAs are installed in the full-height PCIe slot and also require the riser to be installed.

Table 4-31 Storage adapters

Part number	Feature code	Description
Riser card for PCIe adapters		
00FL464	A5JY	nx360 M5 Compute Node Front Riser

Part number	Feature code	Description
Fibre Channel - 16 Gb		
81Y1655	A2W5	Emulex 16Gb FC Single-port HBA for System x
81Y1662	A2W6	Emulex 16Gb FC Dual-port HBA for System x
00Y3337	A3KW	QLogic 16Gb FC Single-port HBA for System x
00Y3341	A3KX	QLogic 16Gb FC Dual-port HBA for System x
Fibre Channel - 8 Gb		
42D0485	3580	Emulex 8Gb FC Single-port HBA for System x
42D0494	3581	Emulex 8Gb FC Dual-port HBA for System x
42D0501	3578	QLogic 8Gb FC Single-port HBA for System x
42D0510	3579	QLogic 8Gb FC Dual-port HBA for System x

For more information, see the list of Product Guides in the Host Bus Adapters category at this website:

<http://lenovopress.com/systemx/hba>

4.1.14 NeXtScale PCIe Native Expansion Tray

The NeXtScale PCIe Native Expansion Tray is a half-wide 1U expansion tray that attaches to the nx360 M5 to provide two full-height full-length double-width PCIe 3.0 x16 slots. The tray supports two GPU adapters or coprocessors.

The use of the PCIe Native Expansion Tray requires that two processors are installed.

Note: The PCIe Native Expansion Tray and the Storage Native Expansion Tray cannot be connected to the same compute node.

Figure 4-11 shows the PCIe Native Expansion Tray attached to an nx360 M5 (shown with the top cover removed) and two NVIDIA GPUs installed.



Figure 4-11 NeXtScale PCIe Native Expansion Tray attached to an nx360 M5 compute node

Ordering information is listed in Table 4-32.

Table 4-32 Ordering information

Part number	Feature code	Description
00Y8393	A4MB	NeXtScale PCIe Native Expansion Tray

When the PCIe Native Expansion Tray is used, it is connected to the compute node through the following riser cards, each providing a PCIe 3.0 x16 connector to the GPUs or coprocessors that are installed in the tray:

- ▶ A 2-slot PCIe 3.0 x24 riser card is installed in the front riser slot (riser slot 1; see Figure 4-3 on page 66). This riser card replaces the standard 1-slot riser that is used to connect standard PCIe cards that are internal to the compute node. The 2-slot riser card offers the following connections:
 - PCIe 3.0 x8 slot for the slot internal to the compute node
 - PCIe 3.0 x16 slots for the front adapter in the PCIe Native Expansion Tray
- ▶ A 1-slot PCIe 3.0 x16 riser card is installed in the rear riser slot (riser slot 2; see Figure 4-3 on page 66). This riser is used to connect the rear adapter in the PCIe Native Expansion Tray.

Only GPUs and coprocessors are supported in the PCIe Native Expansion Tray and only those GPUs and coprocessors that are listed in 4.1.16, “GPU and coprocessor adapters” on page 100. The PCIe Native Expansion Tray also includes the auxiliary power connectors and cables for each adapter slot that is necessary for each supported GPU and coprocessor.

4.1.15 NeXtScale PCIe 2U Native Expansion Tray

The NeXtScale PCIe 2U Native Expansion Tray is a half-wide 2U expansion tray that attaches to the nx360 M5 to provide four full-height full-length double-width PCIe 3.0 x16 slots, 2 at the front and two at the rear of the tray. The tray is designed to support four GPUs or coprocessors, each up to 300 W. In addition, the expansion tray supports up to four 2.5-inch hot-swap SAS/SATA drives with the addition of a hot-swap backplane.

Figure 4-12 shows the PCIe 2U Native Expansion Tray.

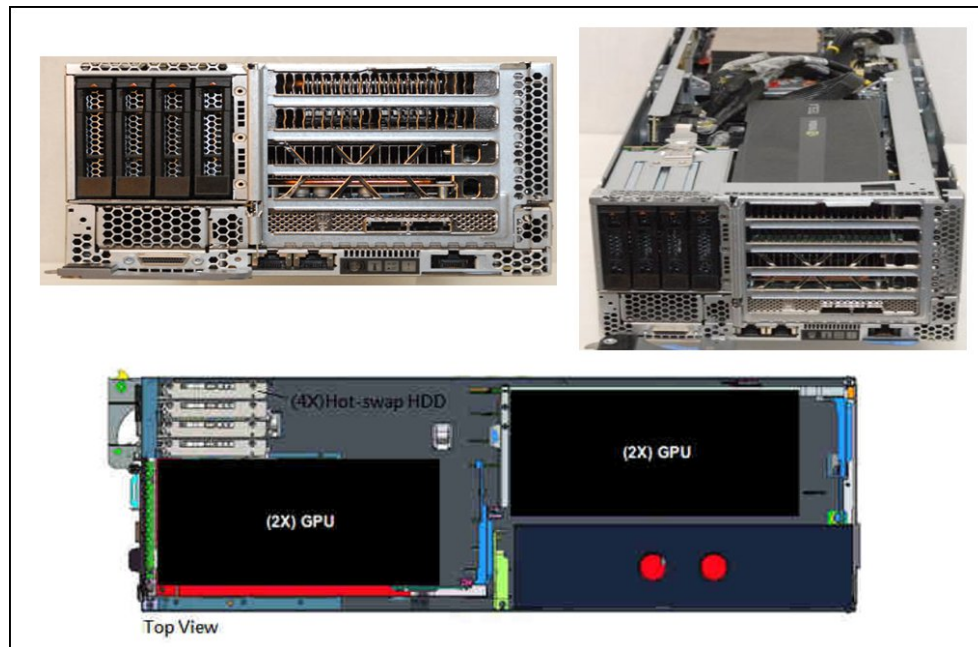


Figure 4-12 PCIe 2U Native Expansion Tray connected to an nx360 M5 server

Ordering information is listed in Table 4-33.

Table 4-33 Ordering information

Part number	Feature code	Description
00MU758	ASYK	NeXtScale PCIe 2U Native Expansion Tray
44X4104	A4A6	4x 2.5" HDD Riser (Backplane and SAS cable)

When the PCIe 2U Native Expansion Tray is used, it is connected to the compute node through two riser cards:

- ▶ The front riser card connects to processor 1 via the PCIe 3.0 x24 slot at the front of the server and provides three slots:
 - One PCIe 3.0 x8 for the full-height half-length slot internal to the nx360 M5 server
 - Two PCIe 3.0 x16 full-height full-length double-width slots for GPUs or coprocessors
- ▶ The rear riser card connects to processor 2 via the PCIe 3.0 x16 slot at the front of the server and provides two slots:
 - Two PCIe 3.0 x16 full-height full-length double-width slots for GPUs or coprocessors

Each riser contains a PEX 8764 PCIe 3.0 switch that enables both x16 slots in the riser to operate at full x16 width. The two riser cards are also connected to each other via two PCIe cables connected to the PCIe switch chips on each riser. The bridge cables allow all four x16 slots to usable even with only 1 processor installed. The connections and slots are shown in Figure 4-13.

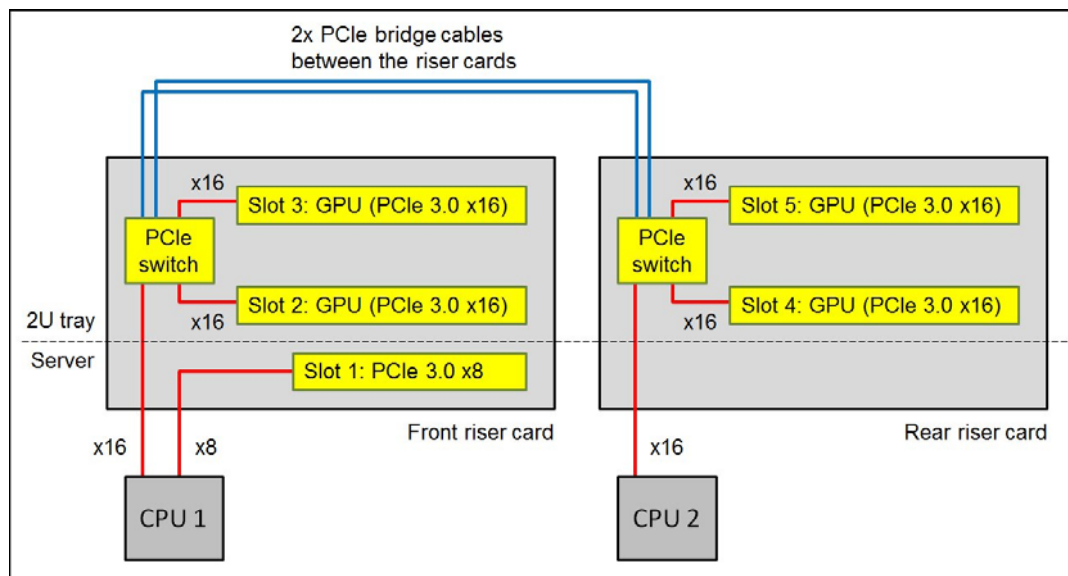


Figure 4-13 Block diagram of the PCIe 2U Native Expansion Tray

Only GPUs and coprocessors are supported in the PCIe 2U Native Expansion Tray and only those that are listed in the following section. The PCIe 2U Native Expansion Tray also includes the auxiliary power connectors and cables for each adapter slot that is necessary for each supported GPU and coprocessor.

4.1.16 GPU and coprocessor adapters

The nx360 M5 supports GPU adapters and coprocessors when the NeXtScale PCIe Native Expansion Tray is attached, as described in the 4.1.14, “NeXtScale PCIe Native Expansion Tray” on page 96. Table 4-34 lists the supported adapters.

Table 4-34 GPU adapters and coprocessors

Part number	Feature code	Description	Power consumption	Maximum supported ^a	
				1U Tray (2 CPUs)	2U Tray (1 CPU / 2 CPUs)
00J6163	A3GQ	Intel Xeon Phi 5110P	225 W	2	No support
00J6162	A3GP	Intel Xeon Phi 7120P	300 W	2	2 / 2
00J6160	A3GM	NVIDIA GRID K1	130W	1	No support
00J6161	A3GN	NVIDIA GRID K2	225 W	2	4 / 4
00D4192	A36S	NVIDIA Tesla K10	225 W	2	No support
00FL133	A564	NVIDIA Tesla K40	235 W	2	4 / 4
00KG133	AS4D	NVIDIA Tesla K80	300 W	2	2 / 2

a. Maximums with PCIe Native Expansion Tray (1U Tray) / PCIe 2U Native Expansion Tray (2U Tray)

The operating systems that are supported by each GPU and coprocessor adapter are listed in 4.1.20, “Supported operating systems” on page 106.

When the PCIe Native Expansion Tray is used, the configuration rules are as follows:

- ▶ Two processors must be installed in the compute node.
- ▶ One or two GPUs or coprocessors can be installed.
- ▶ If two GPU adapters or coprocessors are installed, they must be identical.
- ▶ 1300 W or 1500 W power supplies are required in the chassis.
- ▶ 200 - 240 V AC utility power is required. 100 - 127 V AC is not supported.

When the PCIe 2U Native Expansion Tray is used, the configuration rules are as follows:

- ▶ When 1 processor and 3 or 4 GPUs are installed, the PCIe bridge cable must also be installed
- ▶ When 2 processors and 3 or 4 GPUs are installed, the PCIe bridge cable must be removed
- ▶ All GPU adapters or coprocessors installed in the tray must be identical.
- ▶ 1300 W or 1500 W power supplies are required in the chassis.
- ▶ 200 - 240 V AC utility power is required. 100 - 127 V AC is not supported.

Special bid: Larger numbers of GPUs may be supported via Special Bid than listed in this table, for example up to four NVIDIA K80 or four Intel Xeon Phi 7120P can be supported in the 2U Tray with additional considerations.

These adapters are used for professional graphics to high-performance computing to virtualized and cloud environments. In particular, the following applications are pertinent to NeXtScale System:

- ▶ High-performance computing (HPC): The sheer volume of extra cores that are available in the NVIDIA Tesla GPUs or Intel Xeon Phi can significantly accelerate the millions of calculations that are involved in complex tasks, such as animation rendering or analytic number-crunching. Combining GPUs with CPUs enables so-called GPU-accelerated computing. Highly parallelizable compute-intensive segments of the application are offloaded to GPUs or co-processor while the CPUs run the remainder of the application code. This approach can speed up performance dramatically in scientific, engineering, and enterprise applications.
- ▶ Virtualized and cloud environments: The NVIDIA GRID GPUs enable the following factors:
 - Higher user density (multiple users can share a single GPU)
 - Reduced display latency (the GPU pushes the virtual desktop to the remote user)
 - Greater power efficiency

Table 4-35 lists each of the NVIDIA GPUs that can be used with an nx360 M5 with the attached PCIe Native Expansion Tray.

Table 4-35 Comparison of NVIDIA GPUs

Use Case	High Performance Computing			Virtual Graphics	
Technical Specifications	Tesla K80	Tesla K40	Tesla K10	Grid K2	Grid K1
Peak double-precision FP performance (board)	2.91 Tflops	1.43 Tflops	0.19 Tflops	N/A	N/A
Peak single-precision FP performance (board)	8.74 Tflops	4.29 Tflops	4.58 Tflops	4.30 Tflops	N/A
Number of physical GPUs	2x GK210	1x GK110B	2x GK104		4x GK107
Number of CUDA cores	4992 (2 x 2496)	2,880	3,072 (2 x 1,536)		768 (4 x 192)
Memory size per board	24 GB (2 x 12 GB)	12 GB	8 GB (2 x 4 GB)		16 GB (4 x 4 GB)
Memory bandwidth per board	480 GBps (2 x 240)	288 GBps	320 GBps		116 GBps
Memory I/O	384-bit GDDR5		256-bit GDDR5		128-bit DDR3
Max Power	300W	235W	225W		130W
Aux Power	1 x 8-pin	1 x 8-pin and 1 x 6-pin			1 x 6-pin
PCI Interface	PCIe 3.0 x16				
Physical display connector	None				
Cooling	Passive	Passive, Active	Passive	Passive, Active	Passive
Memory Clock	2.5 GHz	3.0 GHz	2.5 GHz		891 MHz
NVIDIA GPU Boost	Yes		No		
Base Core Clock	562 MHz	745 MHz	745 MHz		850 MHz
Boost Clocks	875 MHz	810 MHz 875 MHz	Not applicable		

Use Case	High Performance Computing			Virtual Graphics	
Technical Specifications	Tesla K80	Tesla K40	Tesla K10	Grid K2	Grid K1
Memory Error Protection	Yes (External & Internal)		Yes (External Only)		No

4.1.17 Integrated virtualization

The server supports VMware vSphere (ESXi), which is installed on a USB memory key. The key is installed in a USB socket inside the server (see Figure 4-3 on page 66 for the USB socket location). Table 4-36 lists the virtualization options.

Table 4-36 Virtualization options: USB Memory Keys

Part number	Feature code	Description	Maximum supported
41Y8298	A2G0	Blank USB Memory Key for VMware ESXi Downloads	1
00ML233	ASN6	USB Memory Key for VMware ESXi 5.1 Update 2	1
00ML235	ASN7	USB Memory Key for VMware ESXi 5.5 Update 2	1

The nx360 M5 also supports the VMware vSphere (ESXi) hypervisor on one or two SD cards with the optional SD Media Adapter for System x. This adapter is installed in a dedicated slot, as shown in Figure 4-3 on page 66.

When only one SD card is installed in the adapter, you can create up to 16 volumes, each of which is presented to UEFI as a bootable device. When two SD Media cards are inserted, volumes can be mirrored (RAID 1) across both cards, up to a total of eight mirrored volumes. The use of mirrored volumes improves system availability because the server remains operational, even if one SD card fails. The RAID functionality is handled internally by the SD Media Adapter.

Table 4-37 lists the available options and how many SD cards are included.

Table 4-37 Virtualization options - SD Cards

Part number	Feature code	Description	SD Cards included
00ML706	A5TJ	SD Media Adapter for Systems x (Option 00ML706 includes 2 blank 32GB SD cards)	2 ^a
00ML700	AS2V	Blank 32GB SD Media for System x	1

a. Option 00ML706 includes two 32 GB SD cards; however, for CTO orders, feature code A5TJ does not include SD media and the SD Cards must be selected separately.

Customized VMware vSphere images can be downloaded from the following website:

http://shop.lenovo.com/us/en/systems/solutions/alliances/vmware/#tab-vmware_vsphere_esxi

4.1.18 Local server management

The nx360 M5 provides local console access through the KVM connector at the front of the server. A console breakout cable is used with this connector, which provides a VGA port, two USB ports, and a DB9 serial port, as shown in Figure 4-14.



Figure 4-14 Console breakout cable

One console breakout cable is shipped with the NeXtScale n1200 enclosure. More cables can be ordered by using the information that is listed in Table 4-38.

Table 4-38 Console breakout cable

Part number	Feature code	Description	Maximum supported
00Y8366	A4AK	Console breakout cable (KVM Dongle cable)	1

Tip: This cable is the same cable that is used with Flex System, but has a different part number because of the included materials.

To aid with problem determination, the server includes light path diagnostics, which is a set of LEDs on the front of the server and inside the server that show you which component is failing. The LEDs are shown in Figure 4-15.

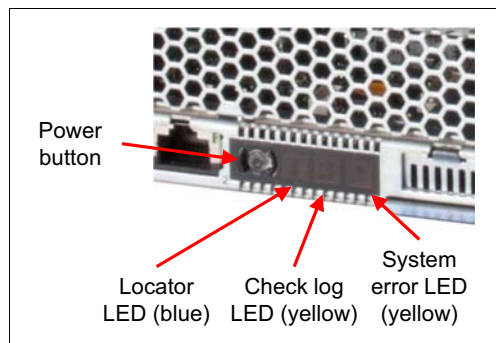


Figure 4-15 Power button and system LEDs

When an error occurs, the system error LED lights up. Review the logs through the interface of the IMMv2 (see 4.1.19, “Remote server management” on page 104). If needed, power off the server and remove it from the enclosure. Then, press and hold the light path diagnostics

button on the system board (see Figure 4-3 on page 66 for the location) to activate the system board LEDs. The LED next to the failed component lights up.

4.1.19 Remote server management

Each NeXtScale nx360 M5 compute node features an Integrated Management Module II (IMM 2.1) onboard and uses the Unified Extensible Firmware Interface (UEFI).

The IMM provides advanced service-processor control, monitoring, and an alerting function. If an environmental condition exceeds a threshold or if a system component fails, the IMM lights LEDs to help you diagnose the problem, records the error in the event log, and alerts you about the problem. Optionally, the IMM also provides a virtual presence capability for remote server management capabilities. The IMM provides remote server management through the following industry-standard interfaces:

- ▶ Intelligent Platform Management Interface (IPMI) version 2.0
- ▶ Simple Network Management Protocol (SNMP) version 3.0
- ▶ Common Information Model (CIM)
- ▶ Web browser

The IMM2.1 also provides the following remote server management capabilities through the **ipmitool** management utility program:

- ▶ Command-line interface (IPMI Shell)

The command-line interface provides direct access to server management functions through the IPMI 2.0 protocol. Use the command-line interface to issue commands to control the server power, view system information, and identify the server. You can also save one or more commands as a text file and run the file as a script.

- ▶ Serial over LAN

Establish a Serial over LAN (SOL) connection to manage servers from a remote location. You can remotely view and change the UEFI settings, restart the server, identify the server, and perform other management functions. Any standard Telnet client application can access the SOL connection.

The NeXtScale nx360 M5 server includes IMM Basic and can be upgraded to IMM Standard and IMM Advanced with Feature on Demand (FoD) licenses.

IMM Basic has the following features:

- ▶ Industry-standard interfaces and protocols
- ▶ Intelligent Platform Management Interface (IPMI) Version 2.0
- ▶ Common Information Model (CIM)
- ▶ Advanced Predictive Failure Analysis (PFA) support
- ▶ Continuous health monitoring
- ▶ Shared Ethernet connection
- ▶ Domain Name System (DNS) server support
- ▶ Dynamic Host Configuration Protocol (DHCP) support
- ▶ Embedded Dynamic System Analysis (DSA)
- ▶ LAN over USB for in-band communications to the IMM
- ▶ SOL
- ▶ Remote power control
- ▶ Server console serial redirection

IMM Standard (as enabled by using the FoD software license key by using part number 90Y3900) has the following features in addition to the IMM Basic features:

- ▶ Remote access through a secure web console
- ▶ Access to server vital product data (VPD)
- ▶ Automatic notification and alerts
- ▶ Continuous health monitoring and control
- ▶ Email alerts
- ▶ Syslog logging support
- ▶ Enhanced user authority levels
- ▶ Event logs that are time stamped, saved on the IMM, and can be attached to email alerts
- ▶ Operating system watchdogs
- ▶ Remote configuration through Advanced Settings Utility (ASU)
- ▶ Remote firmware updating
- ▶ User authentication by using a secure connection to a Lightweight Directory Access Protocol (LDAP) server

IMM Advanced (as enabled by using the FoD software license key by using part number 90Y3901) adds the following features to those features of IMM Standard:

- ▶ Remotely viewing video with graphics resolutions up to 1600 x 1200 at 75 Hz with up to 23 bits per pixel color depths, regardless of the system state
- ▶ Remotely accessing the server by using the keyboard and mouse from a remote client
- ▶ Mapping the CD or DVD drive, diskette drive, and USB flash drive on a remote client, and mapping ISO and diskette image files as virtual drives that are available for use by the server
- ▶ Uploading a diskette image to the IMM memory and mapping it to the server as a virtual drive

The blue-screen capture feature captures the video display contents before the IMM restarts the server when the IMM detects an operating system hang condition. A system administrator can use the blue-screen capture to assist in determining the cause of the hang condition.

Table 4-39 lists the remote management options.

Note: The IMM Advanced upgrade requires the IMM Standard upgrade.

Table 4-39 Remote management options

Part number	Feature codes	Description	Maximum supported
90Y3900	A1MK	Integrated Management Module Standard Upgrade	1
90Y3901	A1ML	Integrated Management Module Advanced Upgrade (requires Standard Upgrade, 90Y3900)	1

The nx360 M5 provides two 1Gbps Ethernet ports standard (see Figure 4-16), one of which (port 1) is configured in UEFI by default to be shared between the operating system and the IMM. In shared mode, this port enables you to connect remotely to the IMM to perform systems management functions. A third Ethernet port is optional (the IMM management Interposer) and provides a dedicated 1Gbps Ethernet connection to the IMM. When the IMM management interposer is in use, port 1 of the two standard Ethernet ports no longer provides access to the IMM but is dedicated to the operating system.

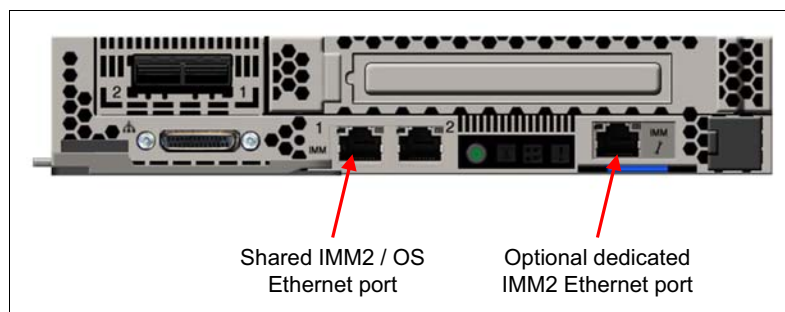


Figure 4-16 IMM ports

Table 4-40 lists the ordering information for the dedicated IMM2 port option.

Table 4-40 Dedicated IMM2 Ethernet port option

Part number	Feature codes	Description	Maximum supported
00FL177	A5JX	nx360 M5 IMM Management Interposer	1

UEFI-compliant server firmware

System x Server Firmware (server firmware) offers several features, including UEFI 2.1 compliance; Active Energy Manager technology; enhanced reliability, availability, and serviceability (RAS) capabilities; and basic input/output system (BIOS) compatibility support. UEFI replaces the BIOS and defines a standard interface between the operating system, platform firmware, and external devices. UEFI-compliant System x servers can boot UEFI-compliant operating systems, BIOS-based operating systems, BIOS-based adapters, and UEFI-compliant adapters.

For more information about the IMM, see the User's Guide that is available at this website:

<http://ibm.com/support/entry/portal/docdisplay?lnocid=migr-5086346>

4.1.20 Supported operating systems

The nx360 M5 server supports the following operating systems:

- ▶ Microsoft Windows Server 2012 R2 (UEFI mode only; Legacy mode not supported)
- ▶ Microsoft Windows Server 2012 (UEFI mode only; Legacy mode not supported)
- ▶ Red Hat Enterprise Linux 7
- ▶ Red Hat Enterprise Linux 6 Server x64 Edition, U5
- ▶ SUSE Enterprise Linux Server (SLES) 12
- ▶ SUSE Linux Enterprise Server 12 with XEN
- ▶ SUSE Linux Enterprise Server 11 for AMD64/EM64T, SP3
- ▶ SUSE Linux Enterprise Server 11 with Xen for AMD64/EM64T, SP3
- ▶ VMware vSphere 6.0 (ESXi)
- ▶ VMware vSphere 5.5 (ESXi), U2

- VMware vSphere 5.1 (ESXi), U2

For more information about the specific versions and service levels that are supported and any other prerequisites, see the following ServerProven website:

<http://www.ibm.com/systems/info/x86servers/serverproven/compat/us/nos/matrix.shtml>

Table 4-41 lists the operating system support for GPUs and coprocessors.

Table 4-41 Operating system support for GPU and coprocessor adapters

Operating system	NVIDIA Tesla K10	NVIDIA Tesla K40m	NVIDIA Tesla K80	NVIDIA Grid K1	NVIDIA Grid K2	Intel Xeon Phi 7120P
Microsoft Windows Server 2012	Y	Y	Y	Y	Y	Y
Microsoft Windows Server 2012 R2	Y	Y	Y	Y	Y	Y
SUSE Linux Enterprise Server 11 for AMD64/EM64T (SP3)	Y	Y	Y	N	N	Y
SUSE Linux Enterprise Server 12	Y	Y	Y	N	N	Y
Red Hat Enterprise Linux 6 Server x64 Edition (U5)	Y	Y	Y	N	N	Y
Red Hat Enterprise Linux 7	Y	Y	Y	N	N	Y
VMware vSphere (ESXi) 5.1	N	N	N	Y	Y	N
VMware vSphere (ESXi) 5.5	N	N	N	Y	Y	N
VMware vSphere (ESXi) 6.0	N	N	N	Y	Y	N

4.1.21 Physical and environmental specifications

The NeXtScale nx360 M5 features the following physical specifications:

- Width: 216 mm (8.5 in.)
- Height: 41 mm (1.6 in.)
- Depth: 659 mm (25.9 in.)
- Maximum weight: 6.17 kg (13.6 lb)

Supported environment

The NeXtScale nx360 M5 compute node complies with ASHRAE class A3 specifications.

The following environment is supported when the node is powered on:

- Temperature: 5 °C - 40 °C (41 °F - 104 °F) up to 950 m (3,117 ft.)
- Above 950 m, de-rated maximum air temperature 1C / 175m
- Humidity, non-condensing: -12 °C dew point (10.4 °F) and 8% - 85% relative humidity
- Maximum dew point: 24 °C (75 °F)
- Maximum altitude: 3050 m (10,000 ft.) and 5 °C - 28 °C (41 °F - 82 °F)

The minimum humidity level for class A3 is the higher (more moisture) of the -12 °C dew point and the 8% relative humidity. These humidity levels intersect at approximately 25 °C. Below

this intersection (~25 °C), the dew point (-12 °C) represents the minimum moisture level, while above it relative humidity (8%) is the minimum.

Moisture levels lower than 0.5 °C DP, but not lower -10 °C DP or 8% relative humidity, can be accepted if appropriate control measures are implemented to limit the generation of static electricity on personnel and equipment in the data center. All personnel and mobile furnishings and equipment must be connected to ground through an appropriate static control system. The following minimum requirements must be met:

- ▶ Conductive materials (conductive flooring, conductive footwear on all personnel that go into the data center, and all mobile furnishings and equipment are made of conductive or static dissipative materials) are used.
- ▶ During maintenance on any hardware, a properly functioning wrist strap must be used by any personnel who comes into contact with IT equipment.

Consider the following points if you adhere to ASHRAE Class A3, Temperature 36 °C - 40 °C (96.8 °F - 104 °F) with relaxed support:

- ▶ A support cloud-like workload with no performance degradation is acceptable (Turbo-Off).
- ▶ Under no circumstance can any combination of worst case workload and configuration result in system shutdown or design exposure at 40 °C.
- ▶ The worst case workload (such as Linpack and Turbo-On) might have performance degradation.

Consider the following specific component restrictions:

- ▶ Processor E5-2699 v3, E5-2697 v3, E5-2667 v3, E5-2643 v3, E5-2637 v3: Temperature: 5 °C - 30 °C (41 °F - 86 °F); Altitude: 0 – 304.8 m (1000 ft.).
- ▶ Intel Xeon Phi 7120P: Temperature: 5 °C - 30 °C (41 °F - 86 °F); Altitude: 0 – 304.8 m (1000 ft.).
- ▶ nx360 M5 with rear HDD: Temperature: 5 °C - 30 °C (41 °F - 86 °F); Altitude: 0 – 304.8 m (1000 ft.).
- ▶ nx360 M5 servers in configurations that also include nx360 M4 servers are not supported by any of the following processors where the TDP is greater than 130 W:
 - E5-2699 v3
 - E5-2697 v3
 - E5-2667 v3
 - E5-2643 v3
 - E5-2637 v3

4.1.22 Regulatory compliance

The server conforms to the following international standards:

- ▶ FCC: Verified to comply with Part 15 of the FCC Rules, Class A
- ▶ Canada ICES-003, issue 5, Class A
- ▶ UL/IEC 60950-1
- ▶ CSA C22.2 No. 60950-1
- ▶ NOM-019
- ▶ Argentina IEC60950-1
- ▶ Japan VCCI, Class A
- ▶ IEC 60950-1 (CB Certificate and CB Test Report)
- ▶ China CCC GB4943.1, GB9254, Class A, and GB17625.1
- ▶ Taiwan BSMI CNS13438, Class A; CNS14336-1

- ▶ Australia/New Zealand AS/NZS CISPR 22, Class A; AS/NZS 60950.1
- ▶ Korea KN22, Class A, KN24
- ▶ Russia/GOST ME01, IEC-60950-1, GOST R 51318.22, and GOST R 51318.24,
- ▶ GOST R 51317.3.2, GOST R 51317.3.3
- ▶ IEC 60950-1 (CB Certificate and CB Test Report)
- ▶ CE Mark (EN55022 Class A, EN60950-1, EN55024, and EN61000-3-2,
- ▶ EN61000-3-3)
- ▶ CISPR 22, Class A
- ▶ TUV-GS (EN60950-1/IEC 60950-1, and EK1-ITB2000)

4.2 NeXtScale nx360 M4 compute node

The NeXtScale nx360 M4 compute node (machine type 5455) is a half-wide, dual-socket server. It supports Intel Xeon E5-2600 v2 series processors up to 12 cores. A total of 12 nx360 M4 servers can be installed into the 6U NeXtScale n1200 enclosure.

This section describes the nx360 M4 compute node and includes the following topics:

- ▶ 4.2.1, “Overview” on page 110
- ▶ 4.2.2, “System architecture” on page 113
- ▶ 4.2.3, “Specifications” on page 115
- ▶ 4.2.4, “Standard models” on page 117
- ▶ 4.2.5, “Processor options” on page 117
- ▶ 4.2.6, “Memory options” on page 118
- ▶ 4.2.7, “Internal disk storage options” on page 123
- ▶ 4.2.8, “NeXtScale Storage Native Expansion Tray” on page 130
- ▶ 4.2.9, “NeXtScale PCIe Native Expansion Tray” on page 133
- ▶ 4.2.10, “GPU and coprocessor adapters” on page 134
- ▶ 4.2.11, “Embedded 1 Gb Ethernet controller” on page 137
- ▶ 4.2.12, “PCI Express I/O adapters” on page 138
- ▶ 4.2.13, “Integrated virtualization” on page 142
- ▶ 4.2.14, “Local server management” on page 143
- ▶ 4.2.15, “Remote server management” on page 144
- ▶ 4.2.16, “External disk storage expansion” on page 146
- ▶ 4.2.17, “Physical specifications” on page 148
- ▶ 4.2.18, “Operating systems support” on page 149

4.2.1 Overview

The NeXtScale nx360 M4 compute node (as shown in Figure 4-17) contains only the essential components in the base architecture to provide a cost-optimized platform. The nx360 M4 compute node provides a dense, flexible solution with a low total cost of ownership.

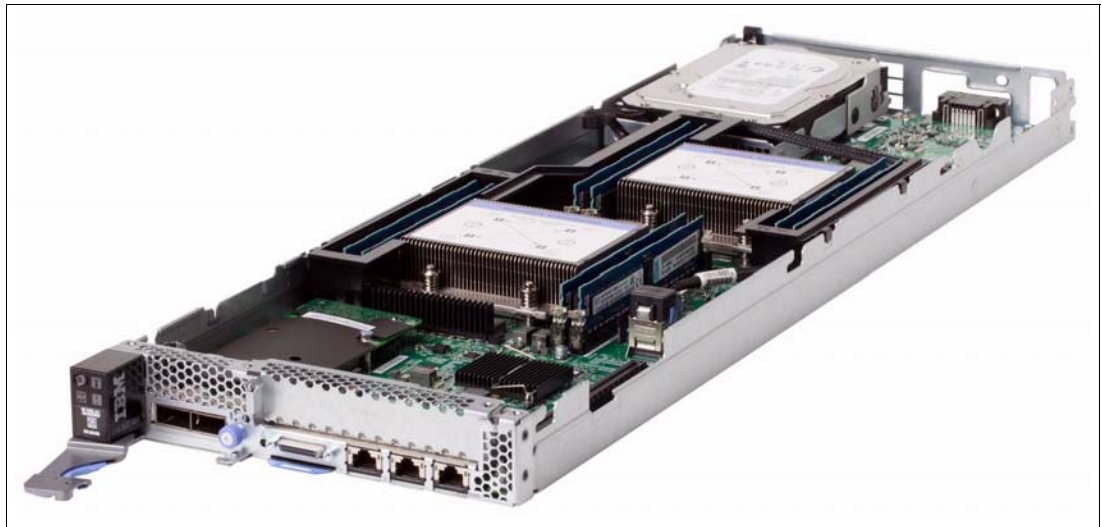


Figure 4-17 NeXtScale nx360 M4 compute node

The nx360 M4 compute nodes fit into the NeXtScale n1200 Enclosure that provides common power and cooling resources. The nx360 M4 compute node can support the following components:

- ▶ One or two Intel Xeon E5-2600 v2 series processors.
- ▶ Up to eight DIMMs of registered DDR3 ECC memory and operating up to 1866 MHz, which provides a total memory capacity of up to 128 GB.
- ▶ One on-board 1 Gb Ethernet port and one on-board 1 Gb Ethernet and management port.
- ▶ An IMM2 port for server remote management and a UEFI, which enables improved setup, configuration, and updates.
- ▶ One 3.5-inch drive bay, or two 2.5-inch drive bays or four 1.8-inch drive bays.
- ▶ Support for more local storage with the use of the Storage Native Expansion Tray. When 4 TB HDDs are used, you can create an ultra-dense storage server with up to 32 TB of total disk capacity within 1U of comparable rack density.
- ▶ A slot for 10 Gb Ethernet or FDR InfiniBand mezzanine for network connectivity without the use of a PCIe slot.
- ▶ PCI Express 3.0 I/O expansion capabilities through a 16x riser cage; more support for two GPUs or coprocessors with the use of the PCIe Native Expansion Tray.

Physical design

Figure 4-18 shows the controls and connections on the front of the server.

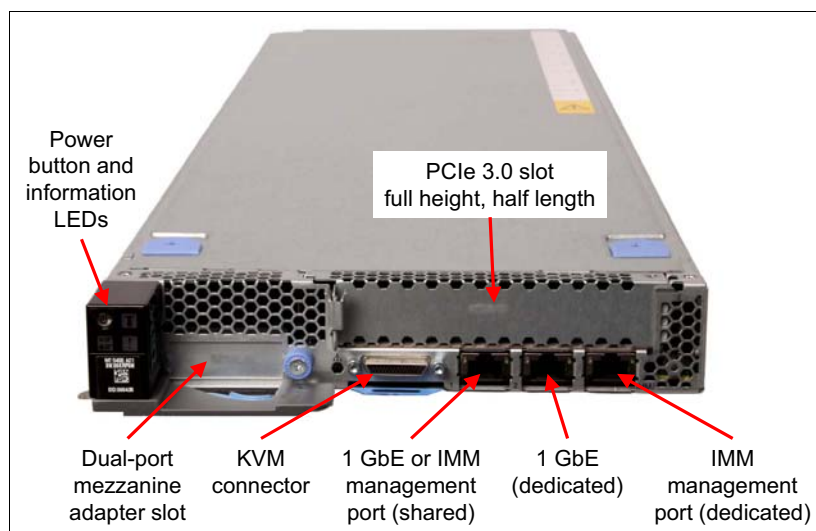


Figure 4-18 Front view of NeXtScale nx360 M4

Figure 4-19 shows the locations of key components inside the server.

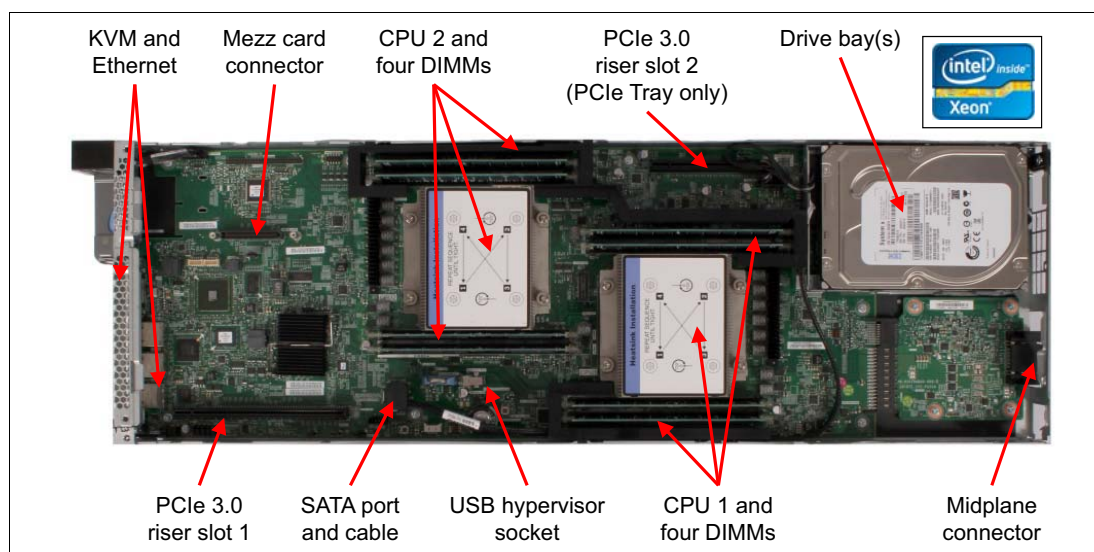


Figure 4-19 Inside view of the NeXtScale nx360 M4

Figure 4-20 shows an exploded window of the platform, in which the major components and options are highlighted.

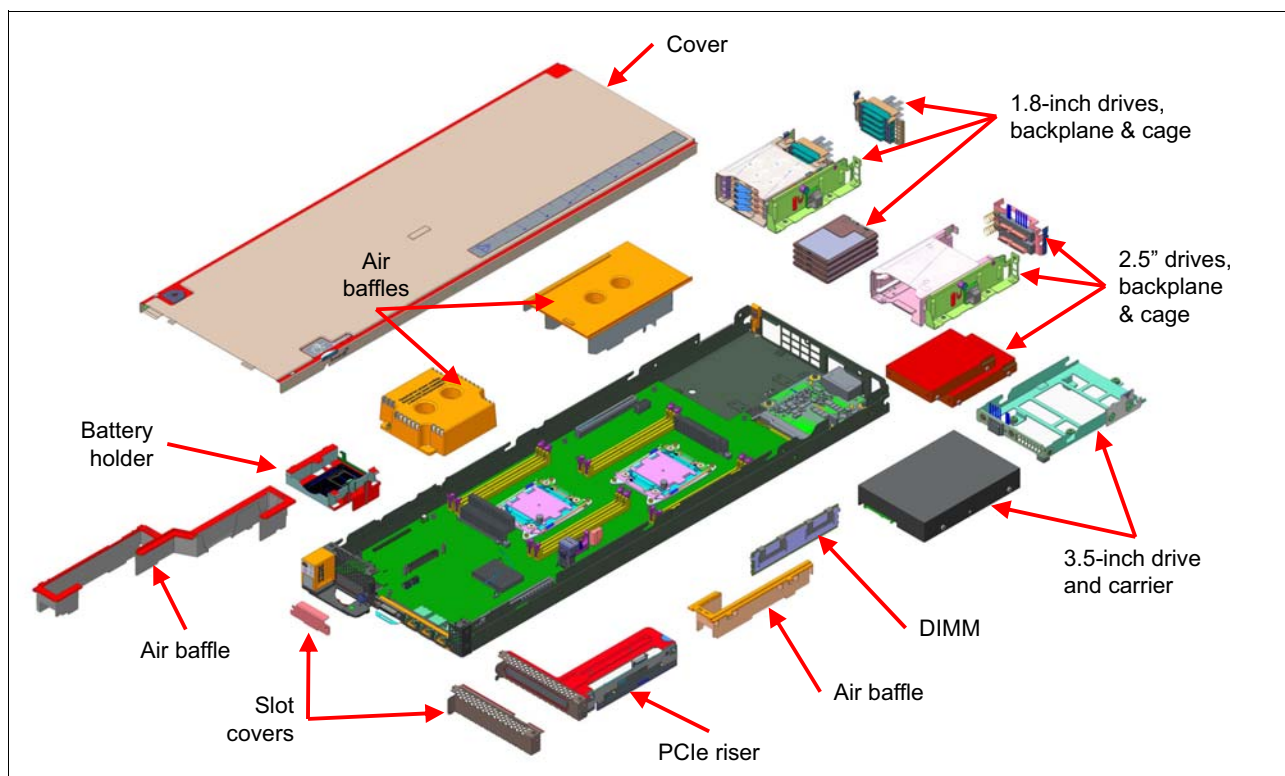


Figure 4-20 Exploded view of the nx360 M4

4.2.2 System architecture

The NeXtScale nx360 M4 compute node features the Intel E5-2600 v2 series processors. The Xeon E5-2600 v2 series processor has models with 6, 8, 10, or 12 cores per processor with up to 24 threads per socket.

The Intel Xeon E5-2600 v2 series processors (formerly known by the Intel code name *Ivy Bridge-EP*) are the successors of the first implementation of Intel's micro architecture that is based on tri-gate transistors, Intel Xeon E5-2600 series (formerly *Sandy Bridge-EP*). The Xeon E5-2600 v2 series uses a 22nm manufacturing process in contrast to the 32nm that was used by its predecessor. By using this new, smaller 22nm tri-gate transistor technology, you can design more powerful processors with better power efficiency.

Such new tri-gate transistor technology enabled a new architecture with which you can share data on-chip through a high-speed ring that is interconnected between all processor cores, the LLC, and the system agent. The system agent houses the memory controller and a PCI Express root complex that provides 40 PCIe 3.0 lanes.

The integrated memory controller in each CPU still supports four memory channels with three DDR3 DIMMs per channel, but now runs at a speed that is up to 1866 MHz. Two QPI links still connect to a second CPU in a dual-socket installation.

The Xeon E5-2600 v2 series is available with up to 12 cores and 30 MB of last-level cache. It features an enhanced instruction set that is called Intel Advanced Vector Extensions (AVX). It doubles the operand size for vector instructions (such as floating-point) to 256 bits and boosts selected applications by up to a factor of two.

The implementation architecture includes Intel Turbo Boost Technology 2.0 and improved power management capabilities. Turbo Boost automatically turns off unused processor cores and increases the clock speed of the cores in use if thermal requirements are still met. Turbo Boost Technology 2.0 uses the integrated design and implements a more granular overclocking in 100 MHz steps instead of 133 MHz steps on older microprocessors.

As with iDataPlex servers, NeXtScale servers support S3 mode. S3 allows systems to come back into full production from low-power state much quicker than a traditional power-on. In fact, cold boot normally takes about 270 seconds; with S3, cold boot occurs in only 45 seconds. When you know that a system is not to be used because of time of day or state of job flow, you can send it into a low-power state to save power and bring it back online quickly when needed.

Table 4-42 lists the differences between Intel's micro architecture implementations. Improvements are highlighted in gray.

Table 4-42 Comparison between Xeon E5-2600 and Xeon E5-2600 v2

	Xeon E5-2600 (Sandy Bridge-EP)	Xeon E5-2600 v2 (Ivy Bridge-EP)
QPI Speed (GT/s)	8.0, 7.2 and 6.4 GT/s	
Addressability	46 bits physical, 48 bits virtual	
Cores	Up to 8	Up to 12
Threads per socket	Up to 16 threads	Up to 24 threads
Last-level Cache (LLC)	Up to 20 MB	Up to 30 MB
Intel Turbo Boost Technology	Yes	

	Xeon E5-2600 (Sandy Bridge-EP)	Xeon E5-2600 v2 (Ivy Bridge-EP)
Memory population	4 channels of up to 3 RDIMMs, 3 LRDIMMs, or 2 UDIMMs	
Maximum memory speed	Up to 1600 MHz	Up to 1866 MHz
Memory RAS features	ECC, Patrol Scrubbing, Sparring, Mirroring, Lockstep Mode, x4/x8 SDDC	
PCIe lanes	40 PCIe 3.0 lanes	
TDP values (W)	130, 115, 96, 80, 70, 60, 50 W	
Idle power targets (W)	15 W or higher 12 W for low-voltage SKUs	10.5 W or higher 7.5 W for low-voltage SKUs

Figure 4-21 shows the NeXtScale nx360 M4 building block.

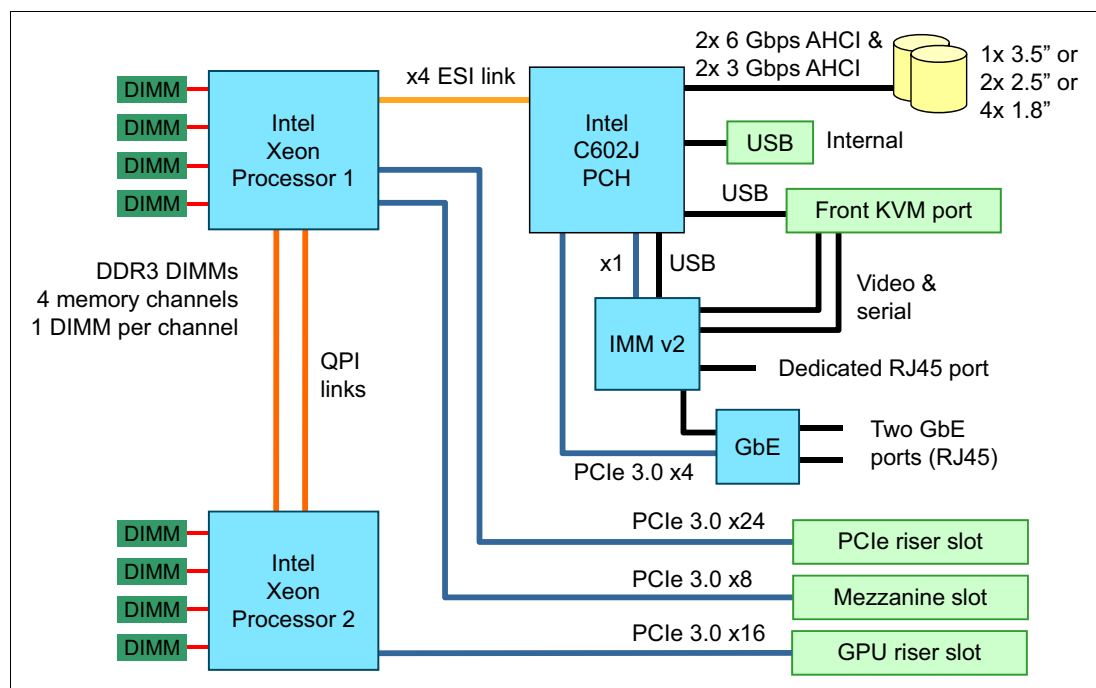


Figure 4-21 NeXtScale nx360 M4 system board block diagram

The nx360 M4 architecture features the following components:

- ▶ Two 2011-pin, Socket R (LGA-2011) processor sockets
- ▶ Intel C602J Platform Controller Hub (PCH)
- ▶ Four memory channels per socket
- ▶ One DIMM per memory channel (DPC)
- ▶ Eight DDR3 DIMM sockets (when only one processor is installed, only four DIMM sockets can be used)
- ▶ Support for UDIMMs and RDIMMs
- ▶ A PCI Express 3.0 x8 slot for Mezzanine card connection that connects to CPU 1
- ▶ A PCI Express 3.0 x24 slot for PCIe riser cage that provides a full-height/half-length (FHHL) slot that connects to CPU 1

- ▶ A PCI Express 3.0 x16 slot that is connected to CPU 2 (for use with attached trays)
- ▶ Dual-port integrated 1 Gb controller
- ▶ An IMMv2 for remote server

Note: Although the Socket R (LGA-2011) processor socket on the nx360 M4 physically fits Xeon E5-2600 series and Xeon E5-2600 v2 series processor, only Xeon E5-2600 v2 is supported as processor options for this platform.

4.2.3 Specifications

Table 4-43 lists the standard specifications of the NeXtScale nx360 M4 compute node.

Table 4-43 Specifications

Components	Specification
Machine type	5455
Firmware	IBM-signed firmware
Form factor	Half-wide, 1U compute node
Supported chassis	NeXtScale n1200 enclosure, 6U high; up to 12 compute nodes per chassis
Processor	Two Intel Xeon Processor E5-2600 V2 series processors; QuickPath Interconnect (QPI) links speed up to 8.0 GT/s. Hyper-Threading Technology and Turbo Boost Technology. Intel C602J (Patsburg-J) chipset. Processor options: <ul style="list-style-type: none"> ▶ 4-core processors up to 3.5 GHz and 15 MB L3 cache ▶ 6-core processors up to 3.5 GHz and 25 MB L3 cache ▶ 8-core processors up to 3.3 GHz and 25 MB L3 cache ▶ 10-core processors up to 3.0 GHz and 25 MB L3 cache ▶ 12-core processors up to 2.7 GHz and 30 MB L3 cache
Memory	Up to 8 DIMM sockets (4 DIMMs per processor) supporting DDR3 DIMMs up to 1866 MHz memory speeds. RDIMMs, UDIMMs and LRDIMMs supported. Four memory channels per processor (one DIMM per channel).
Memory maximum	Up to 256 GB with 8x 32 GB LRDIMMs and two processors.
Memory protection	ECC, memory mirroring, and memory sparing.
Disk drive bays	Inside the nx360 M4: One 3.5-inch simple-swap SATA or two 2.5-inch simple swap SAS/SATA HDDs or SSDs, or four 1.8-inch simple-swap SSDs. Not front accessible. Adding the NeXtScale Storage Native Expansion Tray adds seven 3.5-inch simple-swap drive bays (adds 1U height).
Maximum internal storage	With the Storage Native Expansion Tray: 32 TB that uses 8x 4TB 3.5-inch drives. Without the Storage Native Expansion Tray: 4.0 TB that uses 1x 4TB 3.5-inch drive.
RAID support	On some models: ServeRAID C100 6Gb SATA controller supporting RAID-0, RAID-1 and RAID-10. Implemented in the Intel C600 chipset. Optional hardware RAID with supported 6Gbps RAID controllers.
Optical drive bays	No internal bays; use an external USB drive.
Tape drive bays	No internal bays; use an external USB drive.

Components	Specification
Network interfaces	Two Gigabit Ethernet ports that use onboard Intel I350 Gb Ethernet controller. Optionally, two InfiniBand ports or two 10 GbE ports via a mezzanine card (which does not occupy the available PCIe slot).
PCI Expansion slots	<p>nx360 M4 without PCIe Native Expansion Tray:</p> <ul style="list-style-type: none"> ▶ One PCIe 3.0 x8 mezzanine card slot ▶ One PCIe 3.0 x16 full-height half-length slot <p>nx360 M4 with PCIe Native Expansion Tray (adds 1U height):</p> <ul style="list-style-type: none"> ▶ One PCIe 3.0 x8 mezzanine card slot ▶ One PCIe 3.0 x8 full-height half-length slot ▶ Two PCIe 3.0 x16 full-height full-length double-width slots for GPUs
Ports (server)	Front of the server: KVM connector; with the addition of a console breakout cable (one cable standard with the chassis) supplies one RS232 serial port, one VGA port and two USB ports for local console connectivity. Three 1 Gbps Ethernet ports with RJ45 connectors: one dedicated for systems management (wired to the IMM), one dedicated for use by the operating system and one shared by the IMM and the operating system. One slot for an optional mezzanine card ports (QSFP, SFP+ or RJ45 depending on the card installed). One internal USB port for VMware ESXi hypervisor key.
Ports (chassis)	Rear of the enclosure, provided by the Fan and Power Controller Module for chassis management: Gb Ethernet connection (RJ45) for browser-based remote management, mini-USB serial port for local management.
Cooling	Supplied by the NeXtScale n1200 enclosure. 10 hot-swap dual-rotor 80 mm system fans with tool-less design.
Power supply	Supplied by the NeXtScale n1200 enclosure. Up to six hot-swap power supplies 900 W or 1300 W, depending on the chassis model. Support power policies N+N or N+1 power redundancy and non-redundant power; 80 PLUS Platinum certified.
Video	Matrox G200eR2 video core with 16 MB DDR3 video memory integrated into the IMM2. Maximum resolution is 1600 x 1200 with 16M colors (32 bpp) at 75 Hz, or 1680 x 1050 with 16M colors at 60 Hz. Optional GPUs in PCIe Native Expansion Tray.
Systems management	UEFI, IMM2 with Renesas SH7757 controller, Predictive Failure Analysis, Light Path Diagnostics, Automatic Server Restart, IBM Systems Director and IBM Systems Director Active Energy Manager, ServerGuide. Browser-based chassis management via Ethernet port on the Fan and Power Controller Module on the rear of the enclosure. IMM2 upgrades available to IMM2 Standard and IMM2 Advanced for web GUI and remote presence features.
Security features	Power-on password, administrator's password, Trusted Platform Module 1.2.
Operating systems supported	Red Hat Enterprise Linux, SUSE Linux Enterprise Server, Microsoft Windows Server, VMware vSphere Hypervisor.
Limited warranty	3-year customer-replaceable unit and onsite limited warranty with 9x5/NBD.
Service and support	Optional service upgrades are available through Lenovo Services: 4-hour or 2-hour response time, 8-hour fix time, 1-year or 2-year warranty extension, remote technical support for Lenovo hardware and some Lenovo and OEM software.
Dimensions	NeXtScale nx360 M4 server: Width: 216 mm (8.5 in), height: 41 mm (1.6 in), depth: 659 mm (25.9 in)

Components	Specification
Weight	NeXtScale nx360 M4 maximum weight: 6.05 kg (13.31 lb)

4.2.4 Standard models

Table 4-44 lists the standard models of nx360 M4.

Table 4-44 Standard models of the nx360 M4

Model	Intel Xeon processor ^a	Memory and speed	Disk adapter	Disk bays	Disks	Network
5455-22x	2x Intel Xeon E5-2620 v2 6C 2.1GHz 15MB 1600MHz 80W	2x 4 GB 1600 MHz	6 Gbps SATA (No RAID)	1x 3.5-inch SS bay ^b	Open	2x GbE
5455-42x	2x Intel XeonE5-2660 v2 10C 2.2GHz 25MB 1866MHz 95W	2x 8 GB 1866 MHz	6 Gbps SATA (No RAID)	1x 3.5-inch SS bay	Open	2x GbE
5455-62x	2x Intel XeonE5-2670 v2 10C 2.5GHz 25MB 1866MHz 115W	2x 8 GB 1866 MHz	ServeRAID C100	2x 2.5-inch SS bays	Open	2x GbE

a. Processor detail: Processor quantity and model, cores, core speed, L3 cache, memory speed, and power consumption.

b. SS = simple swap

4.2.5 Processor options

The nx360 M4 supports Xeon E5-2600 v2 processor series. The exact processor options that can be selected for nx360 M4 compute node are listed in Table 4-45.

Table 4-45 Processor options

Part number	Feature code ^a	Intel Xeon processors ^b	Where used
00FL128	A55N / A55W	Intel Xeon E5-2603 v2 4C 1.8GHz 10MB 1333MHz 80W	-
00FL129	A55P / A55X	Intel Xeon E5-2609 v2 4C 2.5GHz 10MB 1333MHz 80W	-
00Y8687	A4MF / A4MK	Intel Xeon E5-2618L v2 6C 2.0GHz 15MB 1333MHz 50W	-
46W2712	A425 / A42F	Intel Xeon E5-2620 v2 6C 2.1GHz 15MB 1600MHz 80W	22x
00FL130	A55Q / A55Y	Intel Xeon E5-2628L v2 8C 2.2GHz 20MB 1600MHz 70W	-
00FL234	A55T / A561	Intel Xeon E5-2630 v2 6C 2.6GHz 15M 1600MHz 80W	-
00FL131	A55R / A55Z	Intel Xeon E5-2630L v2 6C 2.4GHz 15MB 1600MHz 60W	-
00Y8632	A4MD / A4MH	Intel Xeon E5-2637 v2 4C 3.5GHz 15MB 1866MHz 130W	-
46W2719	A42B / A42M	Intel Xeon E5-2640 v2 8C 2.0GHz 20MB 1600MHz 95W	-
00FL126	A55L / A55U	Intel Xeon E5-2643 v2 6C 3.5GHz 25MB 1866MHz 130W	-
00Y8686	A4ME / A4MJ	Intel Xeon E5-2648L v2 10C 2.0GHz 25MB 1866MHz 70W	-
46W2713	A426 / A42G	Intel Xeon E5-2650 v2 8C 2.6GHz 20MB 1866MHz 95W	-

Part number	Feature code ^a	Intel Xeon processors ^b	Where used
00FL132	A55S / A560	Intel Xeon E5-2650L v2 10C 1.7GHz 25M 1600MHz 70W	-
00Y8688	A4MG / A4ML	Intel Xeon E5-2658 v2 10C 2.4GHz 25MB 1866MHz 95W	-
46W2714	A427 / A42H	Intel Xeon E5-2660 v2 10C 2.2GHz 25MB 1866MHz 95W	42x
00FL127	A55M / A55V	Intel Xeon E5-2667 v2 8C 3.3GHz 25MB 1866MHz 130W	-
46W2715	A428 / A42J	Intel Xeon E5-2670 v2 10C 2.5GHz 25MB 1866MHz 115W	62x
46W2716	A429 / A42K	Intel Xeon E5-2680 v2 10C 2.8GHz 25MB 1866MHz 115W	-
46W2717	A42A / A42L	Intel Xeon E5-2690 v2 10C 3.0GHz 25MB 1866MHz 130W	-
46W2720	A42C / A42N	Intel Xeon E5-2695 v2 12C 2.4GHz 30MB 1866MHz 115W	-
46W2721	A42D / A42P	Intel Xeon E5-2697 v2 12C 2.7GHz 30MB 1866MHz 130W	-

a. The first feature code corresponds to the first processor; the second feature code corresponds to the second processor.

b. Processor detail: Model, cores, core speed, L3 cache, memory speed, and TDP power.

Floating point performance: The number of sockets and the processor option that are selected determine the theoretical floating point peak performance, as shown in the following example:

#sockets x #cores per processor x freq x 8 flops per cycle = Gflops

An nx360 M4 compute node with dual socket E5-2680 v2 series 10-core that operates at 2.8 GHz has the following peak performance:

2 x 10 x 2.8 x 8 = 448 Gflops

4.2.6 Memory options

Lenovo DDR3 memory is compatibility tested and tuned for optimal System x performance and throughput. Lenovo memory specifications are integrated into the light path diagnostic tests for immediate system performance feedback and optimum system uptime. From a service and support standpoint, Lenovo memory automatically assumes the system warranty.

The NeXtScale nx360 M4 supports DDR3 memory. The server supports up to four DIMMs when one processor is installed and up to eight DIMMs when two processors are installed. Each processor has four memory channels. There is one DIMM per memory channel (1 DPC).

Figure 4-22 shows the memory channel layout.

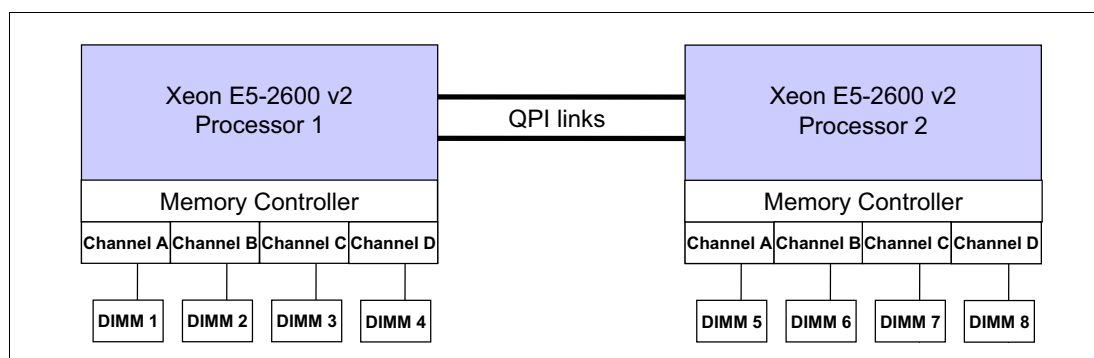


Figure 4-22 Memory channel layout of the NeXtScale nx360 M4

The following rules apply when you select the memory configuration:

- ▶ Each installed processor must have at least one memory DIMM connected.
- ▶ In the nx360 M4, the maximum memory speed of a configuration is the lower of the following two values:
 - The memory speed of the processor (see Table 4-45 on page 117)
 - The memory speed of the DIMM (see Table 4-51 on page 122)
- ▶ The server supports 1.5 V and 1.35 V DIMMs. Mixing 1.5 V and 1.35 V DIMMs in the same server is supported. In such a case, all DIMMs operate at 1.5 V.
- ▶ Mixing UDIMMs and RDIMMs is not supported.
- ▶ Equally distribute DIMMs between sockets for best performance.
- ▶ For optimal performance, populate the four memory channels when one processor is installed and the eight DIMMs when two processors are installed.

Table 4-46 lists the memory options that are available for the nx360 M4 server.

Table 4-46 Memory options

Part number	Feature code	Description	Maximum supported	Models where used
UDIMMs				
00D5012	A3QB	4GB (1x4 GB, 2Rx8, 1.35 V) PC3L-12800 CL11 ECC DDR3 1600MHz LP UDIMM	8	-
RDIMMs - 1866 MHz				
00D5028	A3QF	4GB (1x4 GB, 2Rx8, 1.5 V) PC3-14900 CL13 ECC DDR3 1866MHz LP RDIMM	8	-
00D5040	A3QJ	8GB (1x8 GB, 2Rx8, 1.5 V) PC3-14900 CL13 ECC DDR3 1866MHz LP RDIMM	8	42x, 62x
00D5048	A3QL	16GB (1x16 GB, 2Rx4, 1.5 V) PC3-14900 CL13 ECC DDR3 1866MHz LP RDIMM	8	-
RDIMMs - 1600 MHz				
00D5024	A3QE	4GB (1x4GB, 1Rx4, 1.35V) PC3L-12800 CL11 ECC DDR3 1600MHz LP RDIMM	8	-

Part number	Feature code	Description	Maximum supported	Models where used
46W0735	A3ZD	4GB (1x4 GB, 2Rx8, 1.35 V) PC3-12800 CL13 ECC DDR3 1600MHz LP RDIMM	8	22x
00D5036	A3QH	8GB (1x8GB, 1Rx4, 1.35V) PC3L-12800 CL11 ECC DDR3 1600MHz LP RDIMM	8	-
00D5044	A3QK	8GB (1x8 GB, 2Rx8, 1.35 V) PC3L-12800 CL11 ECC DDR3 1600MHz LP RDIMM	8	-
46W0672	A3QM	16GB (1x16GB, 2Rx4, 1.35V) PC3L-12800 CL11 ECC DDR3 1600MHz LP RDIMM	8	-
LRDIMMs				
46W0761	A47K	32GB (1x32GB, 4Rx4, 1.5V) PC3-14900 CL13 ECC DDR3 1866MHz LP LRDIMM	8	-

The following memory protection technologies are supported:

- ▶ ECC
- ▶ Chipkill (for x4-based memory DIMMs; look for “x4” in the DIMM description; for x8-based memory DIMMs, only ECC protection is supported)
- ▶ Memory mirroring mode
- ▶ Memory lock-step mode

When mirrored mode is used, DIMM 1 and DIMM 2 hold the same data for redundancy purposes. If one DIMM fails, it is disabled and the backup DIMM in the other channel takes over. Likewise, DIMM 3 and DIMM 4 hold the same data and create a mirrored pair. Because memory mirroring is handled in hardware, it is operating system-independent. The total usable memory size is half the total of the installed memory.

Lock-step mode requires paired memory channels (DIMM 1 and DIMM 2, DIMM 3 and DIMM 4) to be populated by the same memory regardless of the size and the organization. Lock-step mode allows Single Device Data Correction (SDDC) memory protection for x8-based memory DIMMs.

DIMM installation order

The NeXtScale nx360 M4 boots with only one memory DIMM installed per processor. However, the suggested memory configuration is to balance the memory across all the memory channels on each processor to use the available memory bandwidth. For best performance, it is recommended to populate all DIMM slots.

The locations of the DIMM sockets relative to the processors are shown in Figure 4-23.

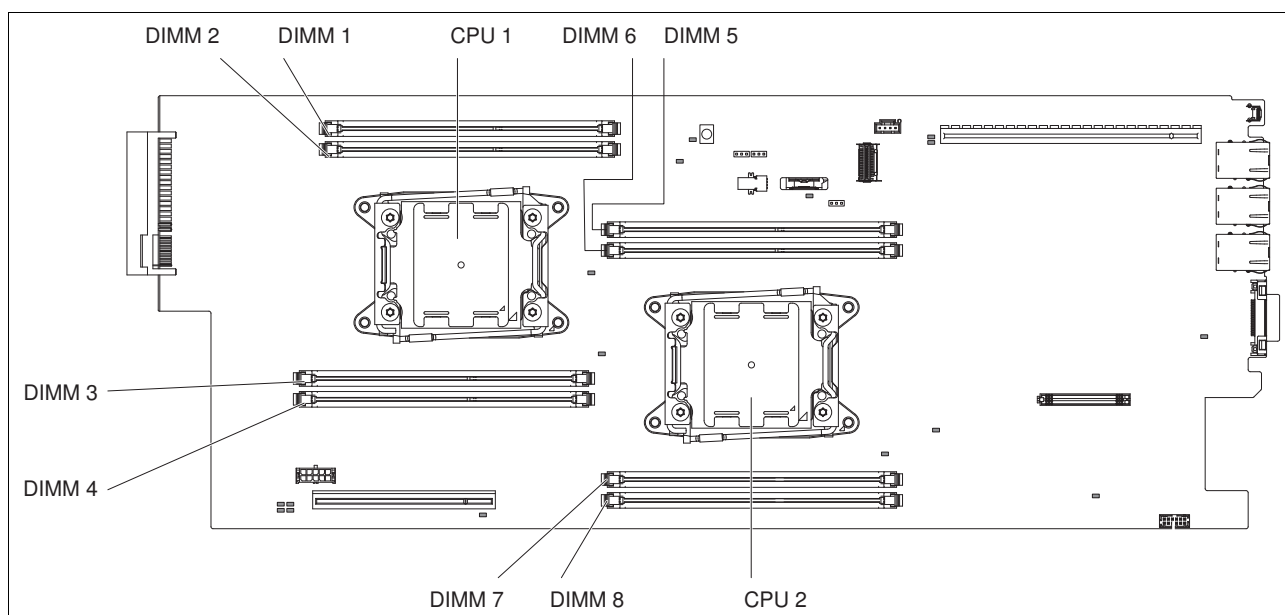


Figure 4-23 Memory DIMM slots

Memory DIMM installation: Independent channel mode

Independent channel mode provides a maximum memory of 64 GB of usable memory with one installed processor, and 128 GB of usable memory with two installed microprocessor (by using 16 GB DIMMs).

Table 4-47 shows DIMM installation if you have one processor that is installed.

Table 4-47 Memory population with one processor installed

	Processor 1			
Number of DIMMs	DIMM 1	DIMM 2	DIMM 3	DIMM 4
1	x			
2	x	x		
3	x	x	x	
4	x	x	x	x

Table 4-48 shows DIMM installation if you have two processors that are installed. A minimum of two memory DIMMs (one for each processor) are required when two processors are installed.

Table 4-48 Memory population table with two processors installed

	Processor 1				Processor 2			
Number of DIMMs	DIMM 1	DIMM 2	DIMM 3	DIMM 4	DIMM 5	DIMM 6	DIMM 7	DIMM 8
2	x				x			
3	x	x			x			

	Processor 1				Processor 2			
4	x	x			x	x		
5	x	x	x		x	x		
6	x	x	x		x	x	x	
7	x	x	x	x	x	x	x	
8	x	x	x	x	x	x	x	x

Memory DIMM installation: Mirrored-channel mode

In mirrored channel mode, the channels are paired and both channels in a pair store the same data.

DIMM 1 and DIMM 2 form a redundant pair, and DIMM 3 and DIMM 4 form the other redundant pair for each microprocessor, as listed in Table 4-49. Because of the redundancy, the effective memory capacity of the compute node is half of the installed memory capacity.

Table 4-49 Memory population with processor stalled: Mirrored channel mode

	Processor 1			
Number of DIMMs	DIMM 1	DIMM 2	DIMM 3	DIMM 4
2	x	x		
4	x	x	x	x

The pair of DIMMs that are installed in each channel must be identical in capacity, type, and rank count.

Table 4-50 lists DIMM installation for memory mirroring if two processors are installed.

Table 4-50 Memory population table with two processors installed: Mirrored channel mode

	Processor 1				Processor 2			
Number of DIMMs	DIMM 1	DIMM 2	DIMM 3	DIMM 4	DIMM 5	DIMM 6	DIMM 7	DIMM 8
4	x	x			x	x		
6	x	x	x	x	x	x		
8	x	x	x	x	x	x	x	x

Table 4-51 (RDIMMs) and Table 4-52 on page 123 (UDIMMs and LRDIMMs) list the maximum memory speeds that are achievable. The tables also list the maximum memory capacity at any speed that is supported by the DIMM and the maximum memory capacity at the rated DIMM speed.

Table 4-51 Maximum memory speeds (RDIMMs)

Spec	RDIMMs	
Rank	Single rank	Dual rank

Spec	RDIMMs		
Part numbers	00D5024 (4 GB) 00D5036 (8 GB)	46W0735 (4 GB) 00D5044 (8 GB) 46W0672 (16 GB)	00D5028 (4 GB) 00D5040 (8 GB) 00D5048 (16 GB)
Rated speed	1600 MHz	1600 MHz	1866 MHz
Rated voltage	1.35 V	1.35 V	1.5 V
Operating voltage	1.35V or 1.5V	1.35V or 1.5V	1.5 V
Max quantity ^a	8	8	8
Largest DIMM	8 GB	16 GB	16 GB
Max memory capacity	64 GB	128 GB	128 GB
Max memory at rated speed	64 GB	128 GB	128 GB
Maximum operating speed (MHz)			
1 DIMM per channel	1600 MHz	1600 MHz	1866 MHz

a. The maximum quantity that is supported is listed for two installed processors. When one processor is installed, the maximum quantity that is supported is half of that shown.

Table 4-52 Maximum memory speeds (UDIMMs and LRDIMMs)

Spec	UDIMMs	LRDIMMs
Rank	Dual rank	Quad rank
Part numbers	00D5012 (4 GB)	46W0761 (32 GB)
Rated speed	1600 MHz	1866 MHz
Rated voltage	1.35V	1.5 V
Operating voltage	1.35 V or 1.5 V	1.5 V
Max quantity ^a	8	8
Largest DIMM	4 GB	32 GB
Max memory capacity	32 GB	256 GB
Max memory at rated speed	32 GB	256 GB
Maximum operating speed (MHz)		
1 DIMM per channel	1600 MHz	1866 MHz

a. The maximum quantity that is supported is shown for two installed processors. When one processor is installed, the maximum quantity that is supported is half of that shown.

4.2.7 Internal disk storage options

This section describes the internal storage options. The RAID controller cards and the conventional and solid-state disks (SSDs) that are supported also are listed.

The NeXtScale nx360 M4 server supports one of the following drive options:

- ▶ One 3.5-inch simple-swap SATA drive
- ▶ Up to two 2.5-inch simple-swap SAS or SATA HDDs or SSDs
- ▶ Up to four 1.8-inch simple-swap SSDs

Consider the following rules for mixing drive types:

- ▶ **Mixing HDDs:** Simple-swap SATA HDDs and simple-swap SAS HDDs can be intermixed in the system, but cannot be intermixed in the same RAID array.
- ▶ **Mixing HDDs and SSDs:** Simple-swap SATA HDDs and simple-swap SAS HDDs can be intermixed with SSDs in the system. SAS or SATA HDDs cannot be configured with SSDs within the same RAID array.

The nx360 M4 also supports seven extra 3.5-inch drive bays if the NeXtScale Storage Native Expansion Tray is attached. The Storage Native Expansion Tray can be used with any of the internal drive configurations to provide the following bay combinations:

- ▶ Eight 3.5-inch simple-swap SATA drives
- ▶ Seven 3.5-inch simple-swap SATA drives and two 2.5-inch simple-swap SATA HDDs
- ▶ Seven 3.5-inch simple-swap SATA drives and four 1.8-inch simple-swap SSDs

For more information about the Native Expansion Tray, see 4.2.8, “NeXtScale Storage Native Expansion Tray” on page 130.

Drive cages for the drives internal to the nx360 M4 are as listed in Table 4-53. Drives that are used in the Storage Native Expansion Tray do not need a cage.

Table 4-53 Drive cages for the drive bay in the nx360 M4

Part number	Feature code	Description	Models where used
None ^a	A41N	nx360 M4 1.8-inch SSD Cage Assembly	-
None ^a	A41K	nx360 M4 2.5-inch HDD Cage Assembly	62x
None ^a	A41J	nx360 M4 3.5-inch HDD Cage Assembly	22x, 42x
00Y8615	A4GE	3.5" HDD RAID cage for nx360 M4 Storage Native Expansion Tray	-

a. CTO only

Figure 4-24 shows the three disk drive bay options that are available for nx360 M4 compute node.

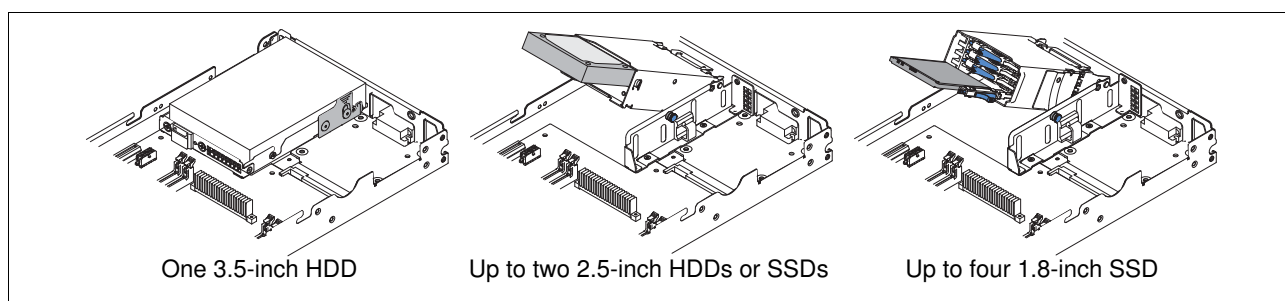


Figure 4-24 Drive bay options

The 3.5-inch drive slides in place. The 2.5-inch and 1.8-inch drive bays pivot up so that you can insert the drives. You then pull the release pin and lower the drive bays back into place.

There are two 3.5-inch drive cages, as listed in the last two rows of Table 4-53. If the Storage Native Expansion Tray is attached to the nx360 M4, the use of the RAID cage (feature A4GE, option 00Y8615) allows you to configure a RAID array that spans all eight drives: the seven in

the storage tray and the 1one drive that is internal to the nx360 M4. Such a configuration is connected to a ServeRAID M1115 adapter or N2115 SAS HBA.

If the 3.5-inch HDD cage (feature A41J) is used, a RAID array can be formed with only the seven drives in the storage tray. In such a configuration, the drives in the storage tray are connected to a ServeRAID M1115 adapter or N2115 SAS HBA. The single drive in the nx360 M4 is connected to the ServeRAID C100.

Controllers for internal storage

The nx360 M4 server supports the following disk controllers:

- ▶ ServeRAID C100: An onboard SATA controller with software RAID capabilities.
- ▶ ServeRAID H1110 SAS/SATA: An entry-level hardware RAID controller that integrates popular SAS technology.
- ▶ ServeRAID M1115 SAS/SATA: An advanced RAID controller with cache and flash modules and energy packs, and software feature upgrades in a flexible offerings structure.
- ▶ N2115 SAS/SATA HBA for System x: A high-performance HBA for internal drive connectivity.

Table 4-54 lists the ordering information for RAID controllers and SAS HBA.

Table 4-54 Controllers for internal storage

Part number	Feature code	Description
RAID controllers		
None	A17T	ServeRAID C100 for System x ^a
81Y4492	A1XL	ServeRAID H1110 SAS/SATA Controller for System x
81Y4448	A1MZ	ServeRAID M1115 SAS/SATA Controller for System x
46C8988	A3MW	N2115 SAS/SATA HBA for System x
Upgrades		
81Y4542	A1X1	ServeRAID M1100 Series Zero Cache/RAID 5 Upgrade (for M1115 only)

a. Windows and Linux only; no support for VMware, Hyper-V, or Xen

ServeRAID C100 controller

The ServeRAID C100 is an integrated SATA controller with software RAID capabilities. It is a cost-effective way to provide reliability, performance, and fault-tolerant disk subsystem management.

The ServeRAID C100 has the following specifications:

- ▶ Supports RAID levels 0, 1, and 10
- ▶ Onboard SATA controller with software RAID capabilities
- ▶ Supports SATA HDDs and SATA SSDs
- ▶ Offers two 6-Gbps SATA ports and two 3-Gbps SATA ports
- ▶ Support for up to two virtual drives
- ▶ Support for virtual drive sizes greater than 2 TB
- ▶ Fixed stripe unit size of 64 KB
- ▶ Support for MegaRAID Storage Manager management software

Note: The ServeRAID C100 is supported by Windows and Linux only. Depending on the operating system version, drivers might need to be downloaded separately. There is no support for VMware, Hyper-V, or Xen

For more information, see the list of Lenovo Press Product Guides in the RAID adapters category at this website:

<http://lenovopress.com/systemx/raid>

ServeRAID H1110 SAS/SATA controller

The ServeRAID H1110 SAS/SATA Controller for System x offers a low-cost, enterprise-grade RAID solution for internal HDDs. It features a PCI Express 2.0 x4 host interface, MD0 form factor, and robust hardware RAID processing engine that is based on the LSI SAS2004 RAID on Chip (ROC) controller.

The ServeRAID H1110 adapter has the following specifications:

- ▶ Four internal 6 Gbps SAS/SATA ports
- ▶ One x4 mini-SAS internal connector (SFF-8087)
- ▶ 6 Gbps throughput per port
- ▶ Based on LSI SAS2004 6 Gbps RAID on Chip (ROC) controller
- ▶ PCIe 2.0 x4 host interface
- ▶ Supports RAID 0, 1, 1E, and 10
- ▶ SAS and SATA drives are supported, but the mixing of SAS and SATA in the same integrated volume is not supported
- ▶ Supports up to two integrated volumes
- ▶ Supports up to two global hot-spare drives
- ▶ Supports drive sizes greater than 2 TB for RAID 0, 1E, and 10 (not RAID 1)
- ▶ Fixed stripe size of 64 KB

For more information, see the list of Lenovo Press Product Guides in the RAID adapters category at this website:

<http://lenovopress.com/systemx/raid>

ServeRAID M1115 controller

The ServeRAID M1115 SAS/SATA Controller is a part of the ServeRAID M Series family that offers a complete server storage solution, which consists of RAID controllers, cache and flash modules, energy packs, and software feature upgrades in an ultra-flexible offerings structure. M1115 also offers a low-cost RAID 0/1/10.

The ServeRAID M1115 adapter has the following specifications:

- ▶ PCI Low Profile, half-length, MD2 form factor
- ▶ A total of eight internal 6 Gbps SAS/SATA ports
- ▶ 6 Gbps throughput per port
- ▶ 533 MHz PowerPC processor with LSI SAS2008 6 Gbps RAID on Chip (ROC) controller
- ▶ PCI Express 2.0 x8 host interface
- ▶ Support for RAID levels 0, 1, and 10 standard
- ▶ Supports RAID levels 5 and 50 with optional M1100 Series RAID 5 upgrade, 81Y4542

- ▶ Support for SAS and SATA HDDs and SSDs
- ▶ Support for intermixing SAS and SATA HDDs and SSDs
- ▶ Support for up to 16 virtual drives, up to 16 drive groups, up to 16 virtual drives per one drive group, and up to 16 physical drives per one drive group
- ▶ Support for virtual drive sizes up to 64 TB
- ▶ Configurable stripe size up to 64 KB
- ▶ Compliant with Disk Data Format (DDF) configuration on disk (COD)
- ▶ S.M.A.R.T. support
- ▶ MegaRAID Storage Manager management software

N2115 SAS/SATA HBA

The N2115 SAS/SATA HBA for System x is an ideal solution for System x servers that require high-speed internal storage connectivity. This eight-port HBA supports direct attachment to SAS and SATA internal HDDs and SSDs. With a low-profile form-factor design, the N2115 SAS/SATA HBA offers two x4 internal mini-SAS connectors.

The N2115 SAS/SATA HBA has the following features and specifications:

- ▶ LSI SAS2308 6 Gbps I/O controller
- ▶ PCI low profile, half-length, kMD2 form factor
- ▶ PCI Express 3.0 x8 host interface
- ▶ Eight internal 6 Gbps SAS/SATA ports (support for 6, 3, or 1.5 Gbps speeds)
- ▶ Up to 6 Gbps throughput per port
- ▶ Two internal x4 Mini-SAS connectors (SFF-8087)
- ▶ Non-RAID (JBOD mode) support for SAS and SATA HDDs and SSDs (RAID not supported)
- ▶ Optimized for SSD performance
- ▶ High-performance IOPS LSI Fusion-MPT architecture
- ▶ Advanced power management support
- ▶ Support for SSP, SMP, STP, and SATA protocols
- ▶ End-to-End CRC with Advanced Error Reporting
- ▶ T-10 Protection Model for early detection of and recovery from data corruption
- ▶ Spread Spectrum Clocking for EMI reductions

For more information, see the list of Lenovo Press Product Guides in the RAID adapters category at this website:

<http://lenovopress.com/systemx/raid>

Table 4-55 lists the adapters that are supported for each drive configuration. SAS HDDs are supported when ServeRAID H1110 or ServeRAID M1115 are included.

Table 4-55 Drive type and RAID adapter support.

Drive type	Quantity of drives supported (max)	Software RAID or no RAID			Hardware RAID with ServeRAID adapter	
		On board SATA	ServeRAID C100	N2115 HBA	With H1110	With M1115
Without Storage Native Expansion Tray attached (one controller)						
1.8-inch SS SATA SSD	4	No	Yes	Yes	Yes	Yes
2.5-inch SS SATA HDD	2	No	Yes	Yes	Yes	Yes
2.5-inch SS SAS HDD	2	No	No	Yes	Yes	Yes
3.5-inch SS SATA HDD	1	Yes	No	No	No	No
With Storage Native Expansion Tray attached (adds 7x 3.5-inch bays) (two controllers) ^a						
1.8-inch SS SATA SSD	4 + 7	No	Yes	Yes	No	Yes
2.5-inch SS SATA HDD	2 + 7	No	Yes	Yes	No	Yes
3.5-inch SS SATA HDD	1 + 7	Yes	Yes	Yes	No	Yes
With Storage Native Expansion Tray attached (adds 7x 3.5-inch bays) (one controller) ^b						
3.5-inch SS SATA HDD	8 ^b	No	No	Yes	No	No

a. The two controllers must be onboard SATA or ServeRAID C100 for internal drives, and N2115 or ServeRAID M115 for drives in the storage tray

b. Requires the 3.5-inch RAID cage (feature A4GE, option 00Y8615)

Using the ServeRAID C100 with 1.8-inch SSDs

The on-board ServeRAID C100 SATA controller provides four channels, two at 3 Gbps and two at 6 Gbps. When a single 3.5-inch SS SATA HDD or two 2.5-inch SS SATA HDDs are installed, drives are connected at 6 Gbps, as shown in Figure 4-25.

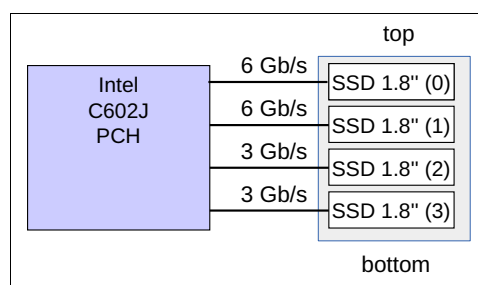


Figure 4-25 Transfer speeds for 1.8-inch drive bays

When you are installing only two 1.8-inch drives, install them in the top two drive bays to maximize performance.

When four 1.8-inch SS SATA SSD drives are installed, the two drives at the top of the cage operate at 6 Gbps while the two at the bottom operate at 3 Gbps. An array that features drives at different speeds performs at the lowest speed. For example, a paired array between top drives operates at 6 Gbps; however, a paired array between a top drive and a bottom drive operates at 3 Gbps.

HDDs and SSDs

Table 4-56 lists the HDD and SSD options that are available for the internal storage of the nx360 M4 server.

Table 4-56 Disk drive options for internal disk storage

Part number	Feature code	Description	Maximum supported
3.5-inch Simple-Swap SATA HDDs			
00AD025	A4GC	4TB 7.2K 6Gbps SATA 3.5-inch HDD for NeXtScale System	1 / 8 ^a
00AD020	A489	3TB 7.2K 6Gbps SATA 3.5-inch HDD for NeXtScale System	1 / 8 ^a
00AD015	A488	2TB 7.2K 6Gbps SATA 3.5-inch HDD for NeXtScale System	1 / 8 ^a
00AD010	A487	1TB 7.2K 6Gbps SATA 3.5-inch HDD for NeXtScale System	1 / 8 ^a
00AD005	A486	500GB 7.2K 6Gbps SATA 3.5-inch HDD for NeXtScale System	1 / 8 ^a
2.5-inch Simple-Swap 10K SAS HDDs			
00FN040	A5NC	1.2TB 10K 6Gbps SAS 2.5-inch HDD for NeXtScale System	2
00AD065	A48G	900GB 10K 6Gbps SAS 2.5-inch HDD for NeXtScale System	2
00AD060	A48F	600GB 10K 6Gbps SAS 2.5-inch HDD for NeXtScale System	2
00AD055	A48D	300GB 10K 6Gbps SAS 2.5-inch HDD for NeXtScale System	2
2.5-inch Simple-Swap 15K SAS HDDs			
00AJ290	A5NG	600GB 15K 6Gbps SAS 2.5-inch HDD for NeXtScale System	2
00AD045	A48E	146GB 15K 6Gbps SAS 2.5-inch HDD for NeXtScale System	2
00AD050	A48H	300GB 15K 6Gbps SAS 2.5-inch HDD for NeXtScale System	2
2.5-inch Simple-Swap SATA HDDs			
00AD030	A48A	250GB 7.2K 6Gbps SATA 2.5-inch HDD for NeXtScale System	2
00AD035	A48B	500GB 7.2K 6Gbps SATA 2.5-inch HDD for NeXtScale System	2
00AD040	A48C	1TB 7.2K 6Gbps SATA 2.5-inch HDD for NeXtScale System	2
2.5-inch simple swap SAS-SSD Hybrid			
00AJ315	A58T	600GB 10K 6Gbps SAS 2.5-inch Hybrid for NeXtScale System	2
2.5-inch simple-swap SSDs			
00FN020	A57K	120GB SATA 2.5-inch MLC Enterprise Value SSD for NeXtScale	2
00FN025	A57L	240GB SATA 2.5-inch MLC Enterprise Value SSD for NeXtScale	2
00FN030	A57M	480GB SATA 2.5-inch MLC Enterprise Value SSD for NeXtScale	2
00FN035	A57N	800GB SATA 2.5-inch MLC Enterprise Value SSD for NeXtScale	2

Part number	Feature code	Description	Maximum supported
1.8-inch simple-swap Enterprise SSDs			
00W1120	A3HQ	100GB SATA 1.8-inch MLC Enterprise SSD	4
1.8-inch simple-swap Enterprise Value SSDs			
00AJ455	A58U	S3500 800GB SATA 1.8-inch MLC Enterprise Value SSD	4
00W1227	A3TH	256GB SATA 1.8-inch MLC Enterprise Value SSD	4
49Y5834	A3AQ	64GB SATA 1.8-inch MLC Enterprise Value SSD	4
00AJ335	A56V	120GB SATA 1.8-inch MLC Enterprise Value SSD	4
00AJ340	A56W	240GB SATA 1.8-inch MLC Enterprise Value SSD	4
00AJ345	A56X	480GB SATA 1.8-inch MLC Enterprise Value SSD	4
00AJ350	A56Y	800GB SATA 1.8-inch MLC Enterprise Value SSD	4

a. One drive is supported without the Storage Native Expansion Tray. Eight drives are supported by the Storage Native Expansion Tray. For more information, see 4.2.8, “NeXtScale Storage Native Expansion Tray” on page 130.

For more information about SSDs, see *Enterprise Solid State Drives for BladeCenter and System x Servers*, TIPS0792, which is available at this website:

<http://lenovopress.com/tips0792>

4.2.8 NeXtScale Storage Native Expansion Tray

The NeXtScale Storage Native Expansion Tray is a half-wide 1U expansion tray that attaches to the nx360 M4 to provide seven extra 3.5-inch simple-swap SATA drives. By using this tray, storage-rich nx360 M4 compute nodes can be configured. By using eight 4 TB drives, such a configuration offers 32 TB of internal direct-attach storage.

Figure 4-26 shows the storage tray attached to an nx360 M4.



Figure 4-26 NeXtScale Storage Native Expansion Tray attached to an nx360 M4 compute node

Ordering information is listed in Table 4-57 on page 131.

Table 4-57 NeXtScale System Internal Storage tray

Part number	Feature code	Description
00Y8546	A4GD	NeXtScale Storage Native Expansion Tray

Figure 4-27 shows the NeXtScale Storage Native Expansion Tray with the cover removed showing seven 3.5-inch drives installed.



Figure 4-27 NeXtScale Storage Native Expansion Tray

When the Storage Native Expansion Tray is used, one of the following disk controller adapters must be installed in the PCIe slot in the nx360 M4:

- ▶ ServeRAID M1115 SAS/SATA Controller for System x, 81Y4448
- ▶ N2115 SAS/SATA HBA for System x, 46C8988

The storage tray connects to both ports on the M1115 controller. Each port supports up to four hard disk drives and both ports are required to support the storage tray.

When the storage tray is installed, the internal 2.5-inch and 1.8-inch disks of the nx360 M4 cannot use the hardware RAID controller. However, they can use the onboard ServeRAID C100 SATA controller to provide software RAID.

The following rules apply to 3.5-inch, 2.5-inch, and 1.8-inch internal drives when the storage tray is installed:

- ▶ If two 2.5-inch SATA drives are installed, the onboard ServeRAID C100 SATA controller is used for RAID with those two drives.
- ▶ If two 2.5-inch SAS drives are installed, the onboard ServeRAID C100 SATA controller cannot be used (the C100 does not support SAS drives) and only operating system software RAID can be used.
- ▶ If four 1.8-inch simple swap SSDs are installed, the onboard ServeRAID C100 SATA controller is used for RAID with those four drives.
- ▶ If one 3.5-inch SATA drive is installed, the following configuration options are available:
 - The single drive in the compute node is connected to the onboard ServeRAID C100 or SATA controller and is independent from the drives in the storage tray. No other components are needed.

- The single drive in the compute node is connected to the ServeRAID M1115 SAS/SATA controller along with the seven drives in the storage tray. All eight drives can be used to form RAID arrays. This configuration requires a special internal storage tray, as listed in Table 4-58.

Table 4-58 NeXtScale 3.5-inch Hardware RAID cage

Part number	Feature code	Description	Maximum supported
00Y8615	A4GE	3.5-inch HDD RAID cage for nx360 M4 Storage Native Expansion Tray	1

Figure 4-28 shows the drive bay numbering. You must populate the drives in the order that is shown.

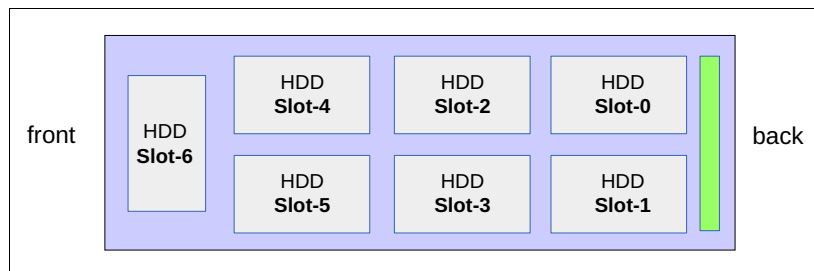


Figure 4-28 Hard disk drive slot numbering

Consider the following rules for installing drives in the storage tray to ensure proper cooling:

- ▶ When an HDD is not present in slots 0, 1, 2, and 3, an HDD filler must be installed to keep the proper air flow. Slots 4, 5, and 6 can be empty.
- ▶ When an HDD fails, it is suggested to keep the failed HDD installed until it is replaced with a new HDD.
- ▶ Some drive bays can be left empty (without a filler). Only bays 0 - 3 require fillers to ensure proper airflow.

Table 4-59 lists the contents of the slots and drive position.

Table 4-59 Drive bay position and fillers required according to the number of drives installed

	7x HDDs	6x HDDs	5x HDDs	4x HDDs	3x HDDs	2x HDDs	1x HDD	No HDD
Bay 0	HDD	HDD	HDD	HDD	HDD	HDD	HDD	Filler
Bay 1	HDD	HDD	HDD	HDD	HDD	HDD	Filler	Filler
Bay 2	HDD	HDD	HDD	HDD	HDD	Filler	Filler	Filler
Bay 3	HDD	HDD	HDD	HDD	Filler	Filler	Filler	Filler
Bay 4	HDD	HDD	HDD	Empty	Empty	Empty	Empty	Empty
Bay 5	HDD	HDD	Empty	Empty	Empty	Empty	Empty	Empty
Bay 6	HDD	Empty	Empty	Empty	Empty	Empty	Empty	Empty

Table 4-60 on page 133 lists 3.5-inch Simple-Swap HDD options for the Storage tray.

Table 4-60 Disk drive options for NeXtScale System Internal Storage tray for nx360

Part number	Feature code	Description	Maximum supported
00AD025	A4GC	4TB 7.2K 6Gbps SATA 3.5-inch HDD for NeXtScale System	7
00AD020	A489	3TB 7.2K 6Gbps SATA 3.5-inch HDD for NeXtScale System	7
00AD015	A488	2TB 7.2K 6Gbps SATA 3.5-inch HDD for NeXtScale System	7
00AD010	A487	1TB 7.2K 6Gbps SATA 3.5-inch HDD for NeXtScale System	7
00AD005	A486	500GB 7.2K 6Gbps SATA 3.5-inch HDD for NeXtScale System	7

4.2.9 NeXtScale PCIe Native Expansion Tray

The NeXtScale PCIe Native Expansion Tray is a half-wide 1U expansion tray that attaches to the nx360 M4 to provide two full-height full-length double-width PCIe 3.0 x16 slots. The tray supports two GPU adapters or coprocessors.

Note: The PCIe Native Expansion Tray and the Storage Native Expansion Tray cannot be connected to the same compute node.

Figure 4-29 shows the PCIe Native Expansion Tray attached to an nx360 M4 (with the top cover removed). Also shown are two NVIDIA GPUs installed.



Figure 4-29 NeXtScale PCIe Native Expansion Tray attached to an nx360 M4 compute node

Ordering information is listed in Table 4-61.

Table 4-61 Ordering information

Part number	Feature code	Description
00Y8393	A4MB	NeXtScale PCIe Native Expansion Tray

When the PCIe Native Expansion Tray is used, it is connected to the compute node via two riser cards, each providing a PCIe x16 connector to the GPUs or coprocessors installed in the tray. The following riser cards are used:

- ▶ A 2-slot PCIe 3.0 x24 riser card is installed in the front riser slot (riser slot 1; see Figure 4-19 on page 112). This riser card replaces the standard 1-slot riser that is used to connect standard PCIe cards internal to the compute node. The 2-slot riser card offers the following connections:
 - PCIe 3.0 x8 slot for the slot internal to the compute node
 - PCIe 3.0 x16 slots for the front adapter in the PCIe Native Expansion Tray
- ▶ A 1-slot PCIe 3.0 x16 riser card is installed in the rear riser slot (riser slot 2; see Figure 4-19 on page 112). This riser is used to connect the rear adapter in the PCIe Native Expansion Tray.

Only GPUs and coprocessors are supported in the PCIe Native Expansion Tray and only those GPUs and coprocessors that are listed in 4.2.10, “GPU and coprocessor adapters” on page 134. The PCIe Native Expansion Tray also includes the auxiliary power connectors and cables for each adapter slot that are necessary for each supported GPU and coprocessor.

4.2.10 GPU and coprocessor adapters

The nx360 M4 supports GPU adapters and coprocessors when the NeXtScale PCIe Native Expansion Tray is attached, as described in 4.2.9, “NeXtScale PCIe Native Expansion Tray” on page 133. Table 4-62 lists the supported adapters.

Table 4-62 GPU adapters and coprocessors

Part number	Feature code	Description	Power consumption	Maximum supported
00J6163	A3GQ	Intel Xeon Phi 5110P	225 W	2
00J6162	A3GP	Intel Xeon Phi 7120P	300 W	2
00J6160	A3GM	NVIDIA GRID K1	130W	2
00J6161	A3GN	NVIDIA GRID K2	225 W	2
00D4192	A36S	NVIDIA Tesla K10	225 W	2
00J6165	A3J8	NVIDIA Tesla K20X	225 W	2
00FL133	A564	NVIDIA Tesla K40	235 W	2

The operating systems that are supported by each GPU and coprocessor adapter is listed in 4.2.18, “Operating systems support” on page 149.

Consider the following configuration rules:

- ▶ The use of GPUs or coprocessors require the use of the NeXtScale PCIe Native Expansion Tray.
- ▶ One or two GPUs or coprocessors can be installed.
- ▶ If two GPU adapters or coprocessors are installed, they must be identical.
- ▶ The 1300 W power supplies are required in the chassis.
- ▶ The 200 - 240 V AC utility power is required; 100-127 V AC is not supported.

For more information about supported quantities of servers with GPUs installed, see to 3.3, “Supported compute nodes” on page 25.

The following advantages of GPUs (and coprocessors) become clear when you compare them to CPUs:

- ▶ CPUs: Fewer but more powerful cores, used primarily for sequential, serial processing.
- ▶ GPUs: More, less powerful cores (up to thousands of them) for simultaneous, parallel processing.
- ▶ Combining GPUs with CPUs enables so-called GPU-accelerated computing (see Figure 4-30). Compute-intensive applications are offloaded to GPUs while the CPUs run the remainder of the application code. This approach can speed up performance dramatically in scientific, engineering, and enterprise applications.

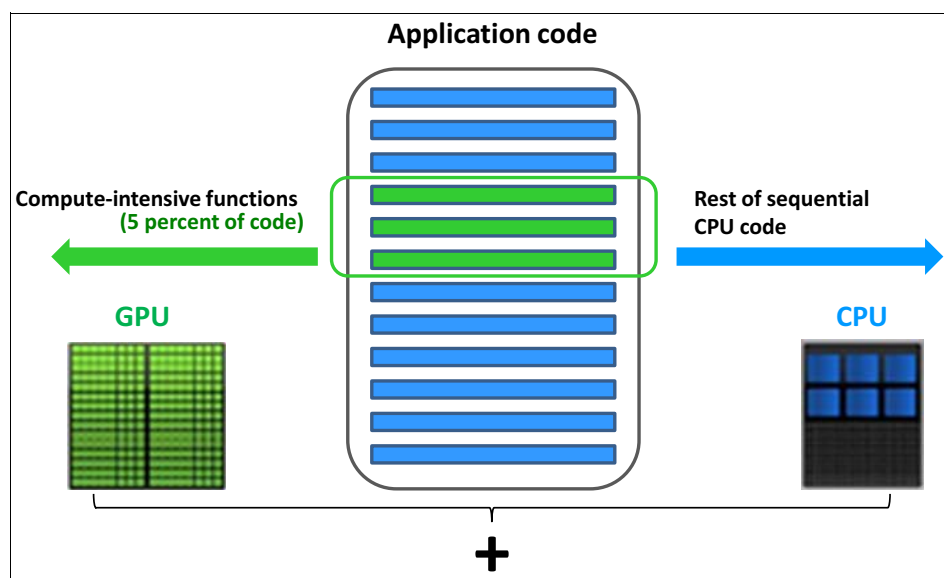


Figure 4-30 GPU-accelerated computing

GPUs are used for everything from consumer gaming and professional graphics to high-performance computing to virtualized and cloud environments. The following applications are pertinent to NeXtScale System:

- ▶ High-performance computing (HPC): The sheer volume of more cores available in GPUs can significantly accelerate the millions of calculations that are involved in complex tasks, such as animation rendering or analytic number-crunching.
- ▶ Virtualized and cloud environments: GPUs enable higher user density (multiple users can share a single GPU), reduced display latency (the GPU pushes the virtual desktop to the remote protocol), and greater power efficiency.

NVIDIA offers several families of GPUs. The following families apply to the nx360 M4:

- ▶ NVIDIA GRID family: Professional-grade virtualized graphics.
NVIDIA GRID GPU cards produce a crisp, responsive, and professional graphics experience. GRID cards expand the advantages of virtualization from the data center to the machine level, with management functions and drivers to allow multiple users and VMs can share GPU resources.

The nx360 M4 supports the following NVIDIA GRID options:

- GRID K2: A dual-slot, 10.5 inch PCI Express Gen3 graphics card with two high-end NVIDIA Kepler GPUs. The GRID K2 features 8 GB of GDDR5 memory (4 GB per GPU), and a 225 W maximum power limit.
- GRID K1: A dual-slot, 10.5 inch PCI Express Gen3 graphics board with four NVIDIA Kepler GPUs. The GRID K1 has 16 GB of DDR3 memory (4 GB per GPU), and a 130 W maximum power limit.

► NVIDIA Tesla family: Scientific-grade computation and analytics

NVIDIA Tesla GPU cards are built to achieve accurate results and withstand long periods of intensive use. The Tesla drivers are coded specifically for scientific-quality computation, and the cards are designed physically and logically for server deployment.

The nx360 M4 supports the following NVIDIA Tesla options:

- Tesla K40: Features 12 GB of memory for demanding HPC and big data analysis. It also includes a Tesla GPUBoost3 feature that can convert power headroom into increased performance.
- Tesla K20X: Designed for double-precision applications across the supercomputing market. The K20X delivers greater than 1.31 TFlops peak double-precision performance.
- Tesla K10: Optimized for single-precision applications. The Tesla K10 combines two ultra-efficient Kepler GPUs to provide high throughput for computations in seismology, signal image processing, and video analytics.

Table 4-63 lists more information each of the NVIDIA GPUs that can be used with an nx360 M4 with the attached PCIe Native Expansion Tray.

Table 4-63 Comparison of NVIDIA GPUs

Use Case	High Performance Computing				Virtual Graphics	
Technical Specifications	Tesla K40	Tesla K20X	Tesla K20	Tesla K10	Grid K2	Grid K1
Peak double-precision FP performance (board)	1.43 Tflops	1.31 Tflops	1.17 Tflops	0.19 Tflops	N/A	N/A
Peak single-precision FP performance (board)	4.29 Tflops	3.95 Tflops	3.52 Tflops	4.58 Tflops	4.30 Tflops (2 x 2.15 Tflops)	N/A
Number of physical GPUs	1x GK110B	1x GK110	1x GK110	2x GK104		4x GK107
Number of CUDA cores	2,880	2,688	2,496	3,072 (2 x 1,536)		768 (4 x 192)
Memory size per board	12 GB	6 GB	5 GB	8 GB (2 x 4 GB)		16 GB (4 x 4 GB)
Memory bandwidth per board	288 GBps	250 GBps	208 GBps	320 GBps		116 GBps
Memory I/O	384-bit GDDR5		320-bit GDDR5	256-bit GDDR5		128-bit DDR3
Max Power	235W	225W				130W

Use Case	High Performance Computing				Virtual Graphics	
Technical Specifications	Tesla K40	Tesla K20X	Tesla K20	Tesla K10	Grid K2	Grid K1
Aux Power	1 x 8-pin and 1 x 6-pin					1 x 6-pin
PCI Interface	Gen 3 x16	Gen 2 x16		Gen3 x16		
Physical display connectors	None					
Cooling	Passive, Active			Passive	Passive, Active	Passive
Memory Clock	3.0 GHz	2.6 GHz		2.5 GHz		891 MHz
NVIDIA GPU Boost	Yes	No				
Base Core Clock	745 MHz	732MHz	706MHz	745 MHz		850 MHz
Boost Clocks	810 MHz 875 MHz	Not applicable				
Memory Error Protection	Yes Enabled (External & Internal)			Yes (External Only)		No

4.2.11 Embedded 1 Gb Ethernet controller

The NeXtScale nx360 M4 includes an embedded 1 Gb Ethernet controller that is built into the system board. It offers 2-Gb Ethernet ports with the following features:

- ▶ Intel I350 Gb Ethernet controller
- ▶ IEEE 802.3 Ethernet interface for 1000BASE-T, 100BASE-TX, and 10BASE-T applications (802.3, 802.3u, and 802.3ab)
- ▶ IPv6 Offloads: Checksum and LSO
- ▶ Wake on LAN support
- ▶ Virtualization: I/OAT, VMDq (eight queues per port), and SR-IOV (PCI SIG compliant)
- ▶ 16 TX and 16 RX queues per port
- ▶ Supports MSI-X
- ▶ Supports SGMI, SCTP, NC-SI
- ▶ Supports IEEE 1588 (TimeSynch) per packet
- ▶ Supports Energy Efficient Ethernet

Figure 4-31 shows the location of the two Ethernet ports and the dedicated IMM2 management port.

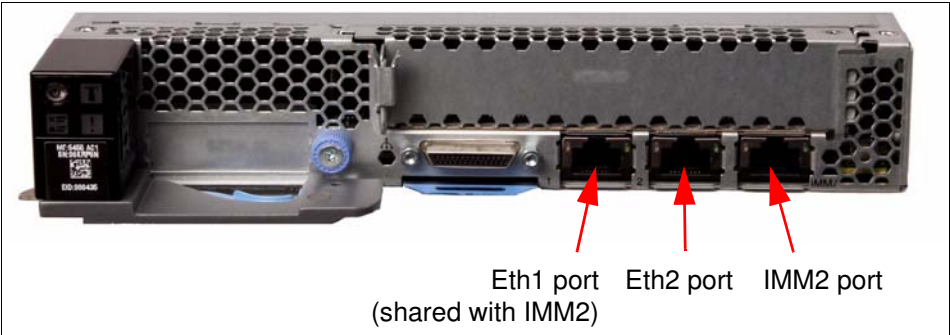


Figure 4-31 Ethernet 1 Gb ports and PCIe I/O slots

4.2.12 PCI Express I/O adapters

The NeXtScale nx360 M4 supports one onboard PCIe card through a mezzanine card and another full-height/half-length through a 1U single-slot riser card.

Mezzanine adapters

The mezzanine card is supported by a dedicated PCIe x8 slot at the front of the server, as shown in Figure 4-32.

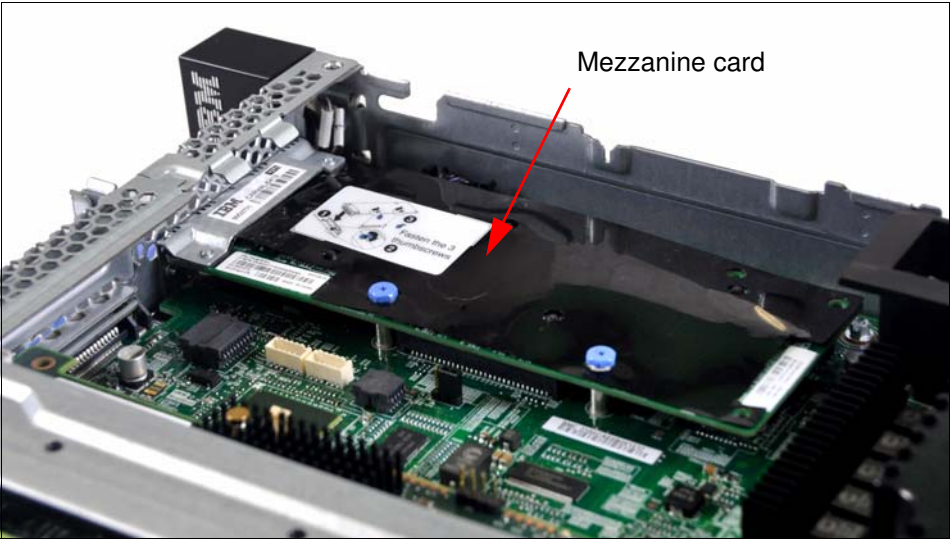


Figure 4-32 Mezzanine card/slot location in NeXtScale nx360 M4

Table 4-64 lists the mezzanine adapters that are supported in nx360 M4.

Table 4-64 Mezzanine adapters

Part number	Feature code	Description
10 Gb Ethernet Mezzanine Card		
00Y7730	A4MC	Emulex Dual Port 10GbE SFP+ Embedded VFA IIIr for System x
49Y7980	A3JS	Intel X520 Dual Port 10GbE SFP+ Embedded Adapter for System x

Part number	Feature code	Description
49Y7990	A3JT	Intel X540 Dual Port 10GBase-T Embedded Adapter for System x
90Y6454	A22H	QLogic Dual Port 10GbE SFP+ Embedded VFA for System x
InfiniBand Mezzanine Card		
00AM476	A4WA	Dual Port FDR10/QDR embedded adapter for nx360 M4
00D4143	A36R	Dual Port FDR Embedded Adapter
FCoE / iSCSI upgrades: Features on Demand		
90Y5179	A2TF	QLogic Embedded VFA FCoE/iSCSI License for System x (FoD) (for 90Y6454)

Single-slot riser card

Extra PCIe cards can be mounted by using a 1U single-slot riser cage. The 1U riser cage provides one PCIe 3.0 x16 slot full-height/half-length. The x16 slot that is provided by the riser (see Table 4-65) can be used to connect a ServeRAID controller or any of the other adapters that are listed in Table 4-66 on page 140 and Table 4-68 on page 141.

PCIe Native Expansion Tray: If the PCIe Native Expansion Tray is selected, this single-slot riser cage is not used. Instead, the tray includes a two-slot riser card. For more information, see 4.2.9, “NeXtScale PCIe Native Expansion Tray” on page 133.

Table 4-65 PCIe riser cage option

Part number	Feature code	Description	Maximum supported
46W2744	A41R	nx360 M4 PCIe riser	1

Figure 4-33 shows the 1U riser cage.

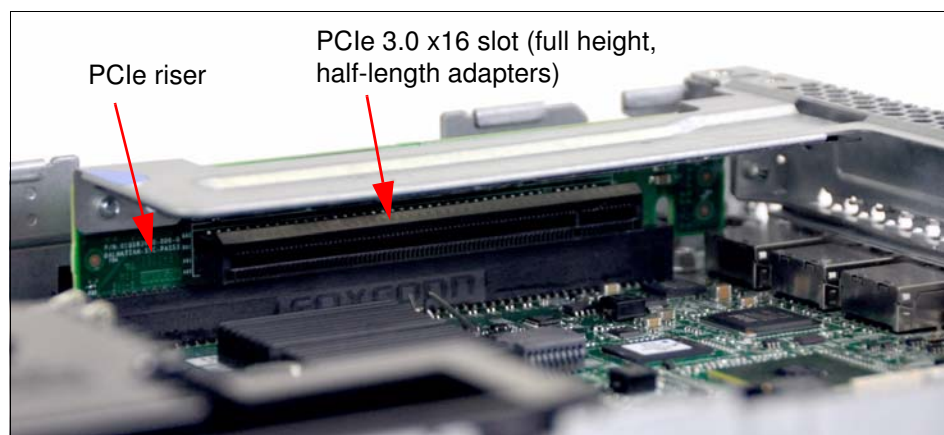


Figure 4-33 1U single slot PCIe riser cage

Network adapters

Table 4-66 on page 140 lists other supported network adapters in the standard full-height half-length PCIe slot.

Table 4-66 Network adapters

Part number	Feature code	Description
40 Gb Ethernet		
00D9550	A3PN	Mellanox ConnectX-3 FDR VPI IB/E Adapter for System x
10 Gb Ethernet		
44T1370	A5GZ	Broadcom NetXtreme 2x10GbE BaseT Adapter for System x
94Y5180	A4Z6	Broadcom NetXtreme Dual Port 10GbE SFP+ Adapter for System x
49Y7910	A18Y	Broadcom NetXtreme II Dual Port 10GBaseT Adapter for System x
00D8540	A4XH	Emulex Dual Port 10GbE SFP+ VFA IIIr for System x
49Y7960	A2EC	Intel X520 Dual Port 10GbE SFP+ Adapter for System x
00D9690	A3PM	Mellanox ConnectX-3 10 GbE Adapter for System x
42C1800	5751	QLogic 10Gb CNA for System x
47C9952	A47H	Solarflare SFN5162F 2x10GbE SFP+ Performant Adapter for System x
47C9960	A47J	Solarflare SFN6122F 2x10GbE SFP+ Onload Adapter for System x
47C9977	A522	Solarflare SFN7122F 2x10GbE SFP+ Flareon Ultra for System x
Gigabit Ethernet		
90Y9352	A2V3	Broadcom NetXtreme I Quad Port GbE Adapter for System x
90Y9370	A2V4	Broadcom NetXtreme I Dual Port GbE Adapter for System x
49Y4230	5767	Intel Ethernet Dual Port Server Adapter I340-T2 for System x
49Y4240	5768	Intel Ethernet Quad Port Server Adapter I340-T4 for System x
00AG500	A56K	Intel I350-F1 1xGbE Fibre Adapter for System x
00AG510	A56L	Intel I350-T2 2xGbE BaseT Adapter for System x
00AG520	A56M	Intel I350-T4 4xGbE BaseT Adapter for System x
InfiniBand		
00D9550	A3PN	Mellanox ConnectX-3 FDR VPI IB/E Adapter for System x

More network adapters are offered as part of the Intelligent Cluster program, as listed in Table 4-67.

Table 4-67 Network adapters that are offered as part of the Intelligent Cluster program

Part number	Feature code	Description
40 Gb Ethernet		
46W0620	A4H5	Chelsio T580-LP-CR Dual-port (QSFP+) 40GbE PCI-E 3.0 Adapter
95Y3459	A2F8	Mellanox ConnectX-3 EN Dual-port QSFP+ 40GbE Adapter
10 Gb Ethernet		
00AE047	A4K1	Mellanox ConnectX-3 EN Single-port SFP+ 10GbE Adapter

Part number	Feature code	Description
00W0053	A2ZQ	Mellanox ConnectX-3 EN Dual-port SFP+ 10GbE Adapter
00Y7006	A3G7	Brocade 1860 Dual-port SFP+ 10GbE Fabric Adapter
00Y7026	A3GE	Brocade 1860 Single-port SFP+ 10GbE Fabric Adapter
46W0609	A4H3	Chelsio T520-LL-CR Dual-port (SFP+) 10GbE PCI-E 3.0 Adapter
46W0615	A4H4	Chelsio T540-CR Quad-port (SFP+) 10GbE PCI-E 3.0 Adapter
InfiniBand		
00W0037	A2YE	Mellanox ConnectX-3 VPI Single-port QSFP FDR IB HCA
00W0041	A2YF	Mellanox ConnectX-3 VPI Dual-port QSFP FDR IB HCA
46W0571	A44E	Mellanox Connect-IB Dual-port QSFP FDR IB PCI-E 3.0 x16 HCA
59Y1888	5763	Intel QLE7340 single-port 4X QDR IB x8 PCI-E 2.0 HCA
95Y3451	A2F6	Mellanox ConnectX-3 VPI Single-port QSFP FDR10 IB HCA
95Y3455	A2F7	Mellanox ConnectX-3 VPI Dual-port QSFP FDR10 IB HCA

Storage host bus adapters

Table 4-68 lists the storage HBAs that are supported by the nx360 M4 server.

Table 4-68 Storage adapters

Part number	Feature code	Description
Fibre Channel: 6 Gb		
81Y1668	A2XU	Brocade 16Gb FC Single-port HBA for System x
81Y1675	A2XV	Brocade 16Gb FC Dual-port HBA for System x
81Y1655	A2W5	Emulex 16Gb FC Single-port HBA for System x
81Y1662	A2W6	Emulex 16Gb FC Dual-port HBA for System x
00Y3337	A3KW	QLogic 16Gb FC Single-port HBA for System x
00Y3341	A3KX	QLogic 16Gb FC Dual-port HBA for System x
Fibre Channel: 8 Gb		
46M6049	3589	Brocade 8Gb FC Single-port HBA for System x
46M6050	3591	Brocade 8Gb FC Dual-port HBA for System x
42D0485	3580	Emulex 8Gb FC Single-port HBA for System x
42D0494	3581	Emulex 8Gb FC Dual-port HBA for System x
42D0501	3578	QLogic 8Gb FC Single-port HBA for System x
42D0510	3579	QLogic 8Gb FC Dual-port HBA for System x
SAS		
46C9010	A3MV	N2125 SAS/SATA HBA for System x

4.2.13 Integrated virtualization

The server supports VMware vSphere (ESXi), which is installed on a USB memory key. The key is installed in a USB socket inside the server. Table 4-69 lists the virtualization options.

Table 4-69 Virtualization options

Part number	Feature code	Description	Maximum supported
41Y8298	A2G0	Blank USB Memory Key for VMware ESXi Downloads	1
41Y8382	A4WZ	USB Memory Key for VMware ESXi 5.1 Update 1	1
41Y8385	A584	USB Memory Key for VMware ESXi 5.5	1

Customized VMware vSphere images can be downloaded from the following website:

<http://ibm.com/systems/x/os/vmware/>

Figure 4-34 shows the USB port inside the nx360 M4 where the USB Memory for VMware ESXi is connected.

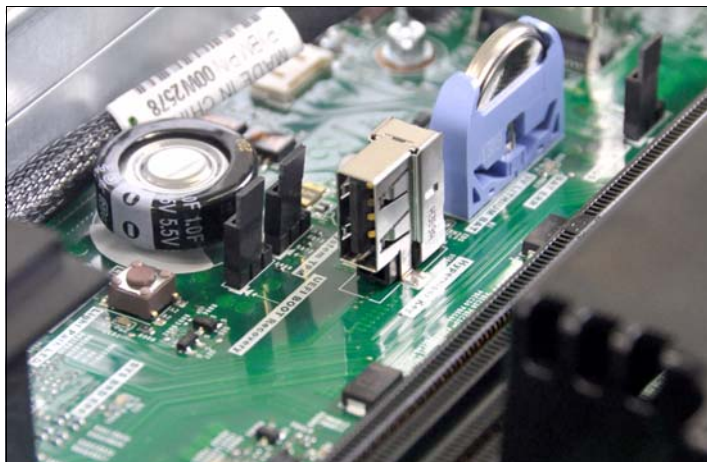


Figure 4-34 Location of the internal USB port for the virtualization key option

4.2.14 Local server management

The nx360 M4 provides local console access through the keyboard, video, and mouse (KVM) connector on the front of the server. A console breakout cable is used with this connector. The cable is shown in Figure 4-35.



Figure 4-35 Console breakout cable

One console breakout cable is shipped with the NeXtScale n1200 Enclosure. Other cables can be ordered. The ordering part number is listed in Table 4-70.

Table 4-70 Console breakout cable

Part number	Feature code	Description	Maximum supported
00Y8366	A4AK	Console breakout cable (KVM Dongle cable)	1

Same cable as Flex System: This cable is the same cable that is used with Flex System, but it has a different part number because of the included materials.

To aid with problem determination, the server includes light path diagnostic indicators, which is a set of LEDs on the front of the server and inside the server that shows you which component is failing. The LEDs are shown in Figure 4-36.

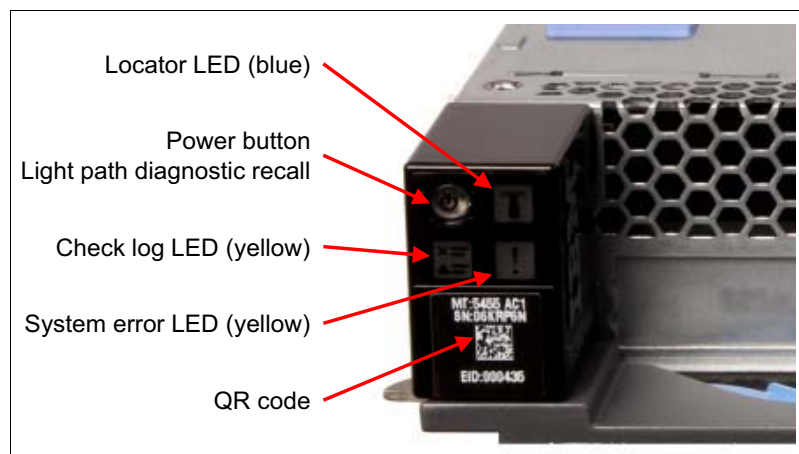


Figure 4-36 NeXtScale nx360 M4 operator panel

The LEDs indicate the following functions:

- ▶ Power button (green): Determines the following power state of the server:
 - Off: AC power is not present.
 - On: Server is powered-on.
 - Flashing rapidly (four times per second): Server is powered-off and is not ready to be turned on.
 - Flashing slowly (once per second): Server is powered-off and is ready to be turned on.
 - Fading on and off: Server is in a reduced power state. To wake up the server, press the power button or use the web interface of the integrated IMM, which is only available with some specific configurations.
- ▶ Locator LED (blue): Location LED controllable by the integrated IMM. This LED is useful to physically identify a system from the management software for maintenance purposes.
- ▶ Check log LED (yellow): Indicates that a condition that causes an event to be logged did occur.
- ▶ System error LED (yellow): Indicates that a fault was detected by the integrated management module (IMM).

When an error occurs, the System error LED lights up. Review the logs through the web interface of the IMMv2 (see 4.2.15, “Remote server management” on page 144). If needed, power off the server and remove it from the enclosure. Then, press and hold the power button to activate the system board LEDs. The LED next to the failed component lights up.

Below the LEDs is a QR code (2D bar code) that, when scanned, contains the machine type and model number (MTM), followed by the serial number of the server.

4.2.15 Remote server management

Each NeXtScale nx360 M4 compute node has an Integrated Management Module II (IMM2) onboard and uses the Unified Extensible Firmware Interface (UEFI) to replace the older interface.

IMM2 provides advanced service-processor control, monitoring, and an alerting function. If an environmental condition exceeds a threshold or if a system component fails, the IMM2 lights LEDs to help you diagnose the problem, records the error in the event log, and alerts you to the problem. The server includes IMM2 Basic and can be upgraded to IMM2 Standard and IMM2 Advanced with FoD licenses; however, you might need to upgrade the system firmware to the latest levels to support these upgrades.

IMM2 Basic has the following features:

- ▶ Industry-standard interfaces and protocols
- ▶ Intelligent Platform Management Interface (IPMI) Version 2.0
- ▶ Common Information Model (CIM)
- ▶ Advanced Predictive Failure Analysis (PFA) support
- ▶ Continuous health monitoring
- ▶ Shared Ethernet connection
- ▶ Domain Name System (DNS) server support
- ▶ Dynamic Host Configuration Protocol (DHCP) support
- ▶ Embedded Dynamic System Analysis (DSA)
- ▶ LAN over USB for in-band communications to the IMM
- ▶ Serial over LAN

- ▶ Server console serial redirection
- ▶ Remote power control

IMM2 Standard (as enabled by using the FoD software license key that uses part number 90Y3900) has the following features in addition to the IMM2 Basic features:

- ▶ Remote access through a secure web console
- ▶ Access to server vital product data (VPD)
- ▶ Automatic notification and alerts
- ▶ Continuous health monitoring and control
- ▶ Email alerts
- ▶ Syslog logging support
- ▶ Enhanced user authority levels
- ▶ Event logs that are time-stamped, saved on the IMM, and can be attached to email alerts
- ▶ Operating system watchdogs
- ▶ Remote configuration through Advanced Settings Utility (ASU)
- ▶ Remote firmware updating
- ▶ User authentication by using a secure connection to a Lightweight Directory Access Protocol (LDAP) server

IMM2 Advanced (as enabled by using the FoD software license key that uses part number 90Y3901) adds the following features in addition to those of IMM Standard:

- ▶ Remotely viewing video with graphics resolutions up to 1600 x 1200 at 75 Hz with up to 23 bits per pixel color depths, regardless of the system state.
- ▶ Remotely accessing the server by using the keyboard and mouse from a remote client.
- ▶ Mapping the CD or DVD drive, diskette drive, and USB flash drive on a remote client, and mapping ISO and diskette image files as virtual drives that are available for use by the server.
- ▶ Uploading a diskette image to the IMM memory and mapping it to the server as a virtual drive.

The blue-screen capture feature captures the video display contents before the IMM restarts the server when the IMM detects an operating-system hang condition. A system administrator can use the blue-screen capture to help determine the cause of the hang condition.

Table 4-71 lists the remote management options.

Note: The IMM2 Advanced upgrade requires the IMM2 Standard upgrade.

Table 4-71 Remote management options

Part number	Feature codes	Description	Maximum supported
90Y3900	A1MK	Integrated Management Module Standard Upgrade	1
90Y3901	A1ML	Integrated Management Module Advanced Upgrade (requires Standard Upgrade, 90Y3900)	1

The nx360 M4 provides a dedicated Ethernet port that allows connection to the IMM2. It is a port to access the IMM2 separately from the onboard two-port 1 Gb Ethernet controller.

Alternatively, the first 1 Gb Ethernet port from the onboard controller can be configured in Shared mode to allow access to the IMM2. With Shared mode enabled, the dedicated IMM port is disabled.

4.2.16 External disk storage expansion

The server supports attachment to external storage expansion enclosures, such as the EXP2500 series, by using the ServeRAID M5120 (6 Gbps) or M5225 (12 Gbps) SAS/SATA Controller. Table 4-72 lists the ordering part numbers.

Table 4-72 RAID controller and features for external locally attached storage

Part number	Feature code	Description
RAID controller		
81Y4478	A1WX	ServeRAID M5120 SAS/SATA Controller for System x
00AE938	A5ND	ServeRAID M5225-2GB SAS/SATA Controller for System x
Hardware upgrades for M5120		
81Y4487	A1J4	ServeRAID M5100 Series 512MB Flash/RAID 5 Upgrade for System x
81Y4559	A1WY	ServeRAID M5100 Series 1GB Flash/RAID 5 Upgrade for System x
Features on Demand Upgrades for M5120		
81Y4546 ^a	A1X3	ServeRAID M5100 Series RAID 6 Upgrade for System x
90Y4273 ^a	A2MC	ServeRAID M5100 Series SSD Performance Key for System x
90Y4318 ^a	A2MD	ServeRAID M5100 Series SSD Caching Enabler for System x

a. Use of any of the FoD upgrades for the M5120 requires one of the hardware upgrades: 81Y4487 or 81Y4559

The ServeRAID M5120 SAS/SATA Controller has the following specifications:

- ▶ Eight external 6 Gbps SAS/SATA ports
- ▶ Up to 6 Gbps throughput per port
- ▶ Two external x4 mini-SAS connectors (SFF-8088)
- ▶ Supports RAID 0, 1, and 10
- ▶ Supports RAID 5 and 50 with optional M5100 Series RAID 5 upgrades
- ▶ Supports RAID 6 and 60 with the optional M5100 Series RAID 6 upgrade
- ▶ Supports 512 MB battery-backed cache or 512 MB or 1 GB flash-backed cache (cache)
- ▶ PCIe 3.0 x8 host interface
- ▶ Based on the LSI SAS2208 6 Gbps ROC controller
- ▶ Supports connectivity to the EXP2512 and EXP2524 storage expansion enclosures

For more information, see *ServeRAID M5120 SAS/SATA Controller*, TIPS0858, which is available at this website:

<http://lenovopress.com/tips0858>

The ServeRAID M5225 SAS/SATA Controller has the following specifications:

- ▶ Eight external 12 Gbps SAS/SATA ports
- ▶ Supports 12, 6, and 3 Gbps SAS and 6 and 3 Gbps SATA data transfer rates
- ▶ Two external x4 mini-SAS HD connectors (SFF-8644)
- ▶ Supports 2 GB flash-backed cache (standard)

- ▶ Supports RAID levels 0, 1, 5, 10, and 50 (standard)
- ▶ Supports RAID 6 and 60 with the optional M5200 Series RAID 6 Upgrade
- ▶ Supports optional M5200 Series Performance Accelerator and SSD Caching upgrades
- ▶ PCIe x8 Gen 3 host interface
- ▶ Based on the LSI SAS3108 12 Gbps ROC controller
- ▶ Supports connectivity to the EXP2512 and EXP2524 storage expansion enclosures

For more information, see *ServeRAID M5225-2GB SAS/SATA Controller*, TIPS1258, which is available at this website:

<http://lenovopress.com/tips1258>

The controllers supports connectivity to the external expansion enclosures that are listed in Table 4-73. Up to nine expansion enclosures can be daisy-chained per one adapter port. For better performance, distribute expansion enclosures evenly across both adapter ports.

Table 4-73 External expansion enclosures

Part number	Description	Maximum supported per one M5120
610012X	EXP2512 Storage Enclosure	17
610024X	EXP2524 Storage Enclosure	9

The external SAS cables that are listed in Table 4-74 support connectivity between external expansion enclosures and the controller.

Table 4-74 External SAS cables for external storage expansion enclosures

Part number	Description	Maximum quantity supported per one enclosure
ServeRAID M5120: Server to Expansion enclosure connectivity (Mini-SAS x4 to Mini-SAS x4)		
39R6531	3 m SAS Cable	1
39R6529	1 m SAS Cable	1
ServeRAID M5225: Server to Expansion enclosure connectivity (Mini-SAS HD x4 to Mini-SAS x4)		
00MJ162	0.6m SAS Cable (mSAS HD to mSAS)	1
00MJ163	1.5m SAS Cable (mSAS HD to mSAS)	1
00MJ166	3m SAS Cable (mSAS HD to mSAS)	1
Expansion enclosure to Expansion enclosure connectivity (Mini-SAS x4 to Mini-SAS x4)		
39R6529	1 m SAS Cable	1
39R6531	3 m SAS Cable	1

Table 4-75 lists the 3.5-inch NL SAS HS HDDs that are supported by EXP2512 external expansion enclosures.

Table 4-75 Drive options for EXP2512 external expansion enclosures

Part number	Description	Maximum quantity supported per one enclosure
00NC555	2TB 7,200 rpm 6Gb SAS NL 3.5-inch HDD	12
00NC557	3TB 7,200 rpm 6Gb SAS NL 3.5-inch HDD	12
00NC559	4TB 7,200 rpm 6Gb SAS NL 3.5-inch HDD	12

Table 4-76 lists the HDDs that are supported by EXP2524 external expansion enclosures.

Table 4-76 Drive options for EXP2524 external expansion enclosures

Part number	Description	Maximum quantity supported per one enclosure
2.5-inch NL SAS HS HDDs		
00NC571	1TB 7,200 rpm 6Gb SAS NL 2.5" HDD	24
2.5-inch SAS HS HDDs		
00NC561	146GB 15,000 rpm 6Gb SAS 2.5" HDD	24
00NC563	300GB 15,000 rpm 6Gb SAS 2.5" HDD	24
00NC565	600GB 10,000 rpm 6Gb SAS 2.5" HDD	24
00NC567	900GB 10,000 rpm 6Gb SAS 2.5" HDD	24
00NC569	1.2TB 10,000 rpm 6Gb SAS 2.5" HDD	24
2.5-inch SAS HS SSDs		
00NC573	200GB 6Gb SAS 2.5" SSD	24
00NC575	400GB 6Gb SAS 2.5" SSD	24

4.2.17 Physical specifications

The NeXtScale nx360 M4 features the following physical specifications:

- ▶ Dimensions:
 - Height: 41 mm (1.6 in)
 - Depth: 658.8 mm (25.9 in)
 - Width: 216 mm (8.5 in)
 - Weight estimation (based on the LFF HDD within computer node): 6.05 kg (13.31 lb)
- ▶ Environment

The NeXtScale nx360 M4 compute node complies with ASHRAE class A3 specifications.
- ▶ Power on (chassis is powered on):
 - Temperature: 5 °C - 40 °C (41 °F - 104 °F)²
 - Humidity, non-condensing: -12 °C dew point (10.4 °F) and 8% - 85% relative humidity^{3, 4}

² A3 - Derate maximum allowable temperature 1°C/175 m above 950 m

- Maximum dew point: 24 °C (75 °F)
- Maximum altitude: 3048 m (10,000 ft.)
- Maximum rate of temperature change: 5 °C per hour (41 °F per hour)⁵
- ▶ Power off⁶:
 - Temperature: 5 °C to 45 °C (41 °F - 113 °F)
 - Relative humidity: 8% - 85%
 - Maximum dew point: 27 °C (80.6 °F)
- ▶ Storage (non-operating):
 - Temperature: 1 °C to 60 °C (33.8 °F - 140 °F)
 - Altitude: 3050 m (10,006 ft.)
 - Relative humidity: 5% - 80%
 - Maximum dew point: 29 °C (84.2 °F)
- ▶ Shipment (non-operating)⁷:
 - Temperature: -40 °C to 60 °C (-40 °F - 140 °F)
 - Altitude: 10,700 m (35,105 ft.)
 - Relative humidity: 5% - 100%
 - Maximum dew point: 29 °C (84.2 °F)⁸
- ▶ Particulate contamination

Airborne particulates and reactive gases that act alone or with other environmental factors, such as, humidity or temperature that might pose a risk to the compute node. For more information about the limits for particulates and gases, see “Particulate contamination” on page 229 in the *NeXtScale nx360 M4 Installation and Service Guide*.

4.2.18 Operating systems support

The nx360 M4 server supports the following operating systems:

- ▶ Microsoft Windows Server 2012 R2
- ▶ Microsoft Windows Server 2012
- ▶ Microsoft Windows Server 2008 R2
- ▶ Microsoft Windows Server 2008 HPC Edition
- ▶ Red Hat Enterprise Linux 7
- ▶ Red Hat Enterprise Linux 6 Server x64 Edition U4
- ▶ Red Hat Enterprise Linux 5 Server x64 Edition U9
- ▶ SUSE Linux Enterprise Server 12
- ▶ SUSE Linux Enterprise Server 11 for AMD64/EM64T SP3

³ The minimum humidity level for class A3 is the higher (more moisture) of the -12 °C dew point and the 8% relative humidity. These levels intersect at approximately 25 °C. Below this intersection (~25 °C), the dew point (-12 °C) represents the minimum moisture level; above the intersection, relative humidity (8%) is the minimum.

⁴ Moisture levels lower than 0.5 °C DP, but not lower -10 °C DP or 8% relative humidity, can be accepted if appropriate control measures are implemented to limit the generation of static electricity on personnel and equipment in the data center. All personnel and mobile furnishings and equipment must be connected to ground via an appropriate static control system. The following items are considered the minimum requirements:

- a. Conductive materials (conductive flooring, conductive footwear on all personnel who go into the datacenter; all mobile furnishings and equipment are made of conductive or static dissipative materials).
- b. During maintenance on any hardware, a properly functioning wrist strap must be used by any personnel who contact IT equipment.

⁵ 5 °C/hr for data centers that are using tape drives and 20 °C/hr for data centers that are using disk drives.

⁶ Chassis is removed from original shipping container and is installed but not in use, for example, during repair, maintenance, or upgrade.

⁷ The equipment acclimation period is 1 hour per 20 °C of temperature change from the shipping environment to the operating environment.

⁸ Condensation (but not rain) is acceptable.

- ▶ VMware vSphere 5.5
- ▶ VMware vSphere 5.1 U1
- ▶ VMware vSphere 5.0 U2

Table 4-77 lists the operating system support for GPUs and coprocessors.

Table 4-77 Operating system support for GPU and coprocessor adapters

Operating system	NVIDIA Tesla K10	NVIDIA Tesla K40m	NVIDIA Tesla K20X	NVIDIA Grid K1	NVIDIA Grid K2	Intel Xeon Phi 7120P	Intel Xeon Phi 5110P
Microsoft Windows Server 2008 R2 (SP1)	Y	Y	Y	Y	Y	Y	Y
Microsoft Windows Server 2008 HPC Edition	Y	Y	Y	N	N	N	N
Microsoft Windows Server 2012	Y	Y	Y	Y	Y	Y	Y
Microsoft Windows Server 2012 R2	Y	Y	Y	Y	Y	Y	Y
SUSE Linux Enterprise Server 11 for AMD64/EM64T (SP3)	Y	Y	Y	N	N	Y	Y
Red Hat Enterprise Linux 5 Server x64 Edition (U10)	Y	Y	Y	N	N	N	N
Red Hat Enterprise Linux 6 Server x64 Edition (U5)	Y	Y	Y	N	N	Y	Y
VMware vSphere (ESXi) 5.0 (U3)	N	N	N	N	N	N	N
VMware vSphere (ESXi) 5.1 (U2)	N	N	N	Y	Y	N	N
VMware vSphere (ESXi) 5.5	N	N	N	Y	Y	N	N

For the more information about the specific versions and service levels that are supported and any other prerequisites, see the following ServerProven website:

<http://www.ibm.com/systems/info/x86servers/serverproven/compat/us/nos/matrix.shtml>

TMTMTM®

Rack planning

A NeXtScale System configuration can consist of many chassis, nodes, switches, cables, and racks. In many cases, it is relevant for planning purposes to think of a system in terms of racks or multiple racks.

In this chapter, we describe best practices for configuring the individual racks. After the rack level design is established, we provide some guidance for designing multiple rack solutions.

This chapter includes the following topics:

- ▶ 5.1, “Power planning” on page 64
- ▶ 5.2, “Cooling” on page 69
- ▶ 5.3, “Density” on page 75
- ▶ 5.4, “Racks” on page 75
- ▶ 5.5, “Cable management” on page 87
- ▶ 5.6, “Rear Door Heat eXchanger” on page 89
- ▶ 5.7, “Top-of-rack switches” on page 93
- ▶ 5.8, “Rack-level networking: Sample configurations” on page 95

5.1 Power planning

In this section, we provide example best practices for configuring power connections and power distribution to meet electrical safety standards while providing a wanted level of redundancy and cooling for racks that contain NeXtScale System chassis and servers.

For more information, see *NeXtScale System Power Requirements Guide*, which is available at this website:

<http://ibm.com/support/entry/portal/docdisplay?lnocid=LNVO-POWINF>

NeXtScale System offers N+1 and N+N power supply redundancy policies at the chassis level. To minimize system cost, it is expected that N+1 or non-redundant power configurations are used; therefore, this use is how the power system was optimized.

When you are planning your power sources for a NeXtScale System rack configuration, consider the following important points:

- ▶ Input voltage

Power supply input is single phase 100 - 120 V, 200 - 240 V alternating current (VAC), or -40.5 - -57 Volt direct current (VDC). For NeXtScale System servers in production environments, 200 - 240 VAC is preferred because it reduces the electrical current requirement. Another consideration is that the power supplies produce less DC output at the lower range.

Restrictions at 110 V: The 900 W CFF power supply is limited to 600 W capacity when operated at low-range voltage (100 - 127 V).

The 1300 W power supply does not support low-range voltages (100 - 127 V).

- ▶ Power distribution unit (PDU) input: single-phase or three-phase power

PDUs can be fed with single-phase or three-phase power. Three-phase power provides more usable power to each PDU and to the equipment. The Lenovo three-phase PDUs separate the phases, which provide single-phase power to the power supplies. The NeXtScale n1200 WCT Enclosure's six power supplies evenly balance the load on three-phase power systems.

- ▶ Single or dual power feed (N+N) to the rack

With a dual-power feed, half of the rack PDUs are powered by one feed and the other half is powered by the second feed, which provides redundant power to the rack.

N+N designs can provide resilience if a power feed must be powered down for maintenance, or if there is a power disruption. However, careful power planning must be done to assure there is adequate power for the NeXtScale systems to keep running on only one power feed.

- ▶ PDU control

PDUs can be switched and monitored, monitored only, or non-monitored. It might be of interest to monitor the power usage at the outlet level of the PDU. More power savings and data center control can be gained with PDUs on which the outlets can be turned on and off.

Single-phase power: In some countries, single-phase power can also be used in such configurations. For more information, see *NeXtScale System Power Requirements Guide*, which is available at this website:

<http://ibm.com/support/entry/portal/docdisplay?lnocid=LNVO-POWINF>

5.1.1 NeXtScale Rack Power Reference Examples

Table 5-1 lists the typical steady-state rack power for common NeXtScale configurations. The power information represents a full rack with 72 x compute nodes and 6 x common switches.

Table 5-1 Typical steady-state power for NeXtScale System rack configurations

Compute Node Configuration	Single Rack Steady-State Power (w/Linpack, Turbo ON/OFF)	Single Rack Steady-State kVA
2x E5-2699 v3 (145 W), 8x 16 GB DDR4 DIMMs, 1x NIC	31.9 kW	32.6 kVA
2x E5-2698 v3 (135 W), 8x 16 GB DDR4 DIMMs, 1x NIC	30.4 kW	31 kVA
2x E5-2695 v3 (120 W), 8x 16 GB DDR4 DIMMs, 1x NIC	28.1 kW	28.7 kVA
2x E5-2660 v3 (105 W), 8x 16 GB DDR4 DIMMs, 1x NIC	25.9 kW	26.4 kVA
2x E5-2640 v3 (90 W), 8x 16 GB DDR4 DIMMs, 1x NIC	23.6 kW	24.1 kVA
2x E5-2630 v3 (85 W), 8x 16 GB DDR4 DIMMs, 1x NIC	22.8 kW	23.3 kVA
2x E5-2650L v3 (65 W), 8x 16 GB DDR4 DIMMs, 1x NIC	19.8 kW	20.2 kVA
2x E5-2630L v3 (55 W), 8x 16 GB DDR4 DIMMs, 1x NIC	18.3 kW	18.7 kVA

5.1.2 Examples

Data center power can be supplied from a single power feed, which can be protected by a facility UPS. The power cabling that is shown in Figure 5-1 uses three PDUs. The PDUs can be supplied by using a 60 A, 200 - 240 V three-phase source or a 32 A, 380 - 415 V three-phase source. The colors (red, blue, green) show the phases as separated by the PDUs.

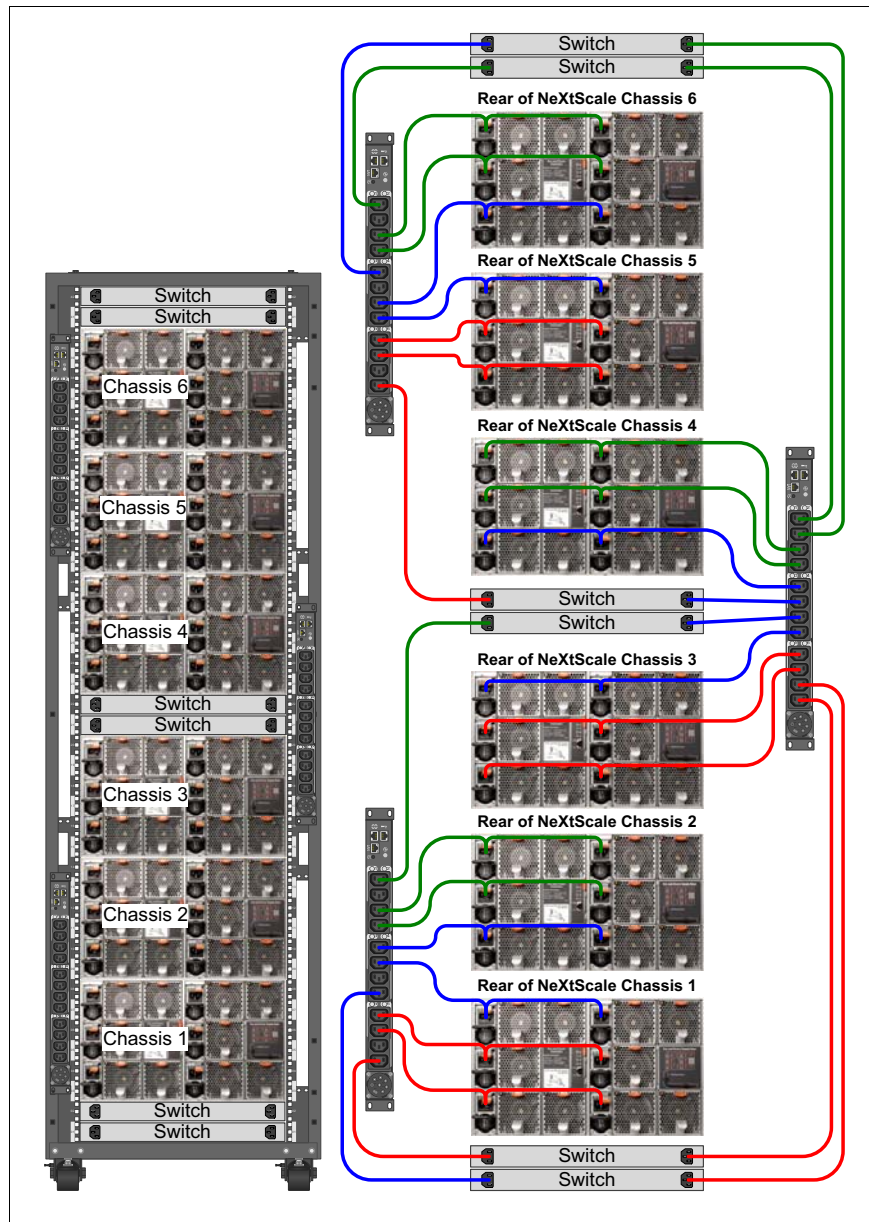


Figure 5-1 Six chassis and six switches that are connected to three PDUs

The PDUs that are shown are in vertical pockets in the rear of an 42U 1100mm Enterprise V2 Dynamic Rack. For more information about the rack, see 5.4.1, “Rack Weight” on page 75.

This configuration requires “Y” power cables for the chassis power supplies. Part numbers are listed in Table 5-2.

Table 5-2 Y cable part numbers

Part number	Description
00Y3046	1.345 m, 2x C13 to C14 Jumper Cord, Rack Power Cable
00Y3047	2.054 m, 2x C13 to C14 Jumper Cord, Rack Power Cable

Figure 5-2 shows the connections from four 1U PDUs. Each PDU has 12 outlets; therefore, 48 outlets are available. There are six NeXtScale n1200 Enclosures, each with six power supplies, so there are 36 power supplies to be connected. In all, there are 12 outlets that are not connected to chassis.

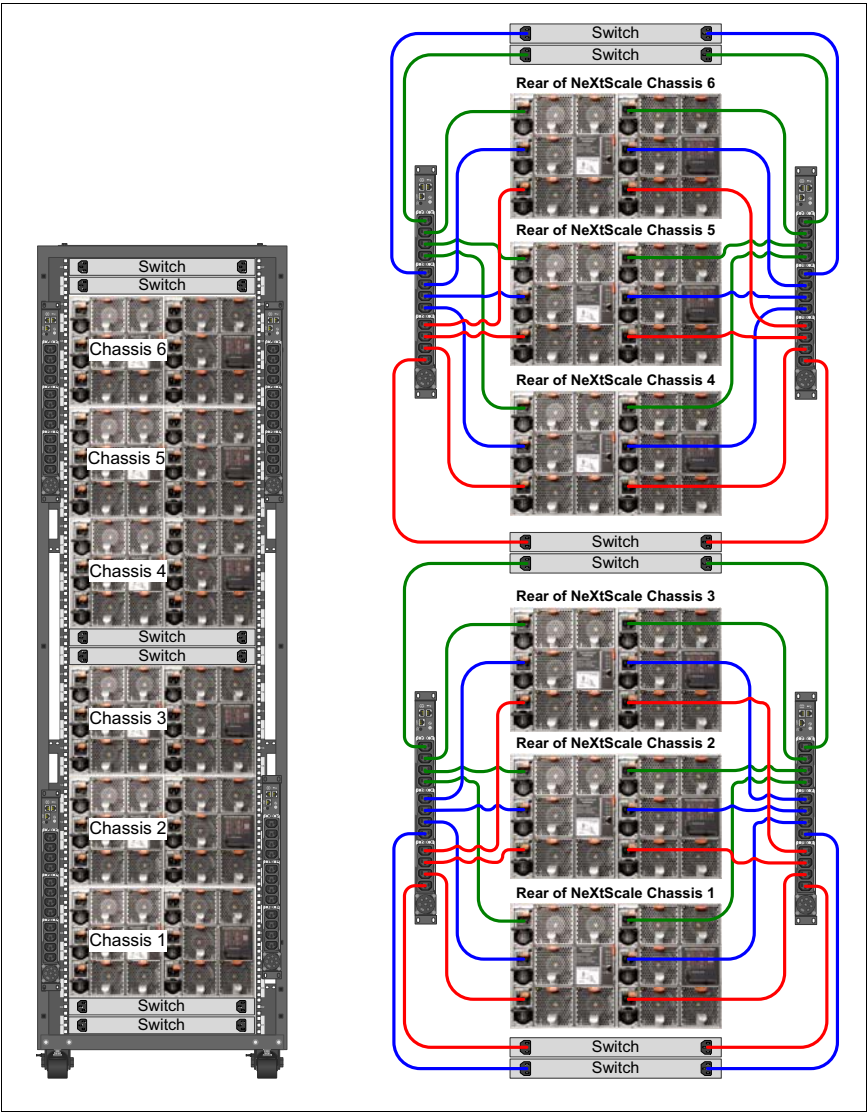


Figure 5-2 Sample power cabling: Six chassis and six switches

With 6U of rack space left open, it is possible to put six 1U devices in the rack and provide two independent power connections to each device. This configuration can provide power

redundancy to the chassis, servers, and optional devices that are installed in the rack, depending on the specifications of the equipment.

5.1.3 PDUs

There are several power distribution units that can be used with NeXtScale System. Listed in Table 5-3 are 1U rack units, which have 12 C13 outlets and are supplied with 200 - 240 V 60 A, three-phase power or 380 - 415 V 32 A, three-phase power.

0U PDUs: The Lenovo 0U PDU should not be used in the 42U 1100mm Enterprise V2 Dynamic Rack with the NeXtScale n1200 Enclosure because there is inadequate clearance between the rear of the chassis and the PDU.

Table 5-3 PDUs for use with NeXtScale System

Part number	Description
Switched and Monitored PDU	
46M4005	1U 12 C13 Switched and Monitored 60 A 3-Phase PDU
46M4004	1U 12 C13 Switched and Monitored PDU without line cord
Monitored PDU	
39M2816	DPI C13 Enterprise PDU without line cord
Basic PDU	
39Y8941	Enterprise C13 PDU

The switched and monitored PDU (part number 46M4005) includes an attached line cord with IEC 609 3P+G plug. The other PDUs require a line cord, which is listed in Table 5-4.

Table 5-4 Line cord part number

Part number	Description
40K9611	32 A 380-415V IEC 309 3P+N+G (non-US) Line Cord

For more information about selecting appropriate PDUs to configure, see the Lenovo Power Guide, which is available at this website:

<http://ibm.com/support/entry/portal/docdisplay?lnocid=LNVO-POWINF>

For more information about PDUs, see the Lenovo Press Product Guides that are available at this website:

<http://lenovopress.com/systemx/power>

5.1.4 UPS units

There are several rack-mounted UPS units that can be used with the NeXtScale systems, which are listed in Table 5-5. In larger configurations, UPS service is often supplied at the data center level.

Table 5-5 UPS units for use with NeXtScale System

Part number	Description
55945KX	RT5kVA 3U Rack UPS (200-240Vac)
55946KX	RT6kVA 3U Rack UPS (200-240Vac)
55948KX	RT8kVA 6U Rack UPS (200-240Vac)
55949KX	RT11kVA 6U Rack UPS (200-240Vac)
55948PX	RT8kVA 6U 3:1 Phase Rack UPS (380-415Vac)
55949PX	RT11kVA 6U 3:1 Phase Rack UPS (380-415Vac)

For more information about UPS units, see the Lenovo Press Product Guides that are available at this website:

<http://lenovopress.com/systemx/power>

5.2 Cooling

Cooling is an important consideration in designing any server solution. It can require careful planning for large-scale environments. After the power planning is complete, calculating the amount of heat to be dissipated is relatively straightforward. For each W of power that is used, 3.414 British Thermal Units per hour (BTU/hr) of cooling is required.

5.2.1 Planning for air cooling

It is often more difficult to determine the volume of air, which is commonly measured in cubic feet per minute (CFM), that is required to cool the servers. The following factors influence the air volume requirement of the servers:

- ▶ The intake air temperature, which affects how much air is required to cool the components to an acceptable temperature.
- ▶ Humidity, which affects the air's thermal conductivity.
- ▶ Altitude and barometric pressure, which affect air density.
- ▶ Airflow impedance in the environment.

Table 5-6 shows typical and maximum airflow values for the NeXtScale n1200 Enclosure at 25C inlet ambient temperature.

Table 5-6 Typical and maximum airflow for the NeXtScale n1200 Enclosure

Typical Airflow (CFM)	Maximum Airflow (CFM)
225 CFM	600 CFM

In data centers that contain several power dense racks, extracting the used warm air and providing enough chilled air to the equipment intakes can be challenging.

In these environments, one of the primary considerations is preventing warm air from recirculating directly from the equipment exhaust into the cold air intake. It is important to use filler panels to block any unused rack space. Part numbers for kits of five filler panels, which install quickly and without tools, are listed in Table 5-7.

Table 5-7 Blank filler panel kits: Five panels per kit

Part number	Description
25R5559	1U Quick Install Filler Panel Kit (quantity five)
25R5560	3U Quick Install Filler Panel Kit (quantity five)

The blank filler panel kits that are listed in Table 5-7 can be used next to the NeXtScale n1200 Enclosure. However, next to switches that are mounted in the front of the rack requires the use of the 1U Pass Through Bracket (part number 00Y3011), as described in 5.4.5, “Rack options” on page 83. This bracket is required if there is an empty 1U space above or below a front-mounted switch to prevent air recirculation from the rear to the front of the switch. Front-mounted 1U switches that are used with NeXtScale are recessed 75 mm behind the front rack mounting brackets to provide sufficient room for cables. A standard filler panel does not contact a switch that is recessed 75 mm; therefore, hot air recirculation can occur.

The next areas to watch for air recirculation are above, below, and around the sides of the racks. Avoiding recirculation above the racks can be difficult to address because this area is typically the return path to the air conditioner and often there is overhead lighting, plumbing, cable trays, or other things with which to contend.

Examples of the use of recirculation prevention plates or stabilizer brackets to prevent under rack air return, and joining racks to restrict airflow around the sides is described in 5.4.1, “Rack Weight” on page 75. We also suggest covering any unused space in rack cable openings.

There are various approaches to addressing the cooling challenge. The use of traditional computer room air conditioners with a raised floor and perforated floor tiles requires numerous floor space that is dedicated to perforated tiles and generous space for air return paths.

Smaller rooms are possible with air curtains, cold air and warm air separation ducting, fans or blowers, or other airflow modification schemes. These approaches typically create restricted spaces for airflow, which are windy, noisy, inflexible in their layout, and difficult to move around in.

Another approach is to contain a single rack or a row of racks in purpose built enclosure with its own air movement and cooling equipment. When multiplied by several racks or rows of racks, this approach is expensive and less space efficient. It also often requires the rack or row to be powered off if the enclosure must be opened for maintenance of the equipment or the enclosure.

Lenovo suggests the use of Rear Door Heat Exchanger, as described in 5.6, “Rear Door Heat eXchanger” on page 89. This option is a relatively low-cost, low-complexity, space, and power efficient solution to the cooling challenge.

5.3 Density

With the NeXtScale System, most data centers should sustain a density increase of approximately 28% as compared to iDataPlex (Lenovo's previous high-density server design) even with 6U in each 42U rack that is reserved for switches or cabling. These server types are compared in Figure 5-6 and use an area 10 x 10 floor tile-wide deep-rack. NeXtScale servers can easily provide double the density of 1U rack server configurations.

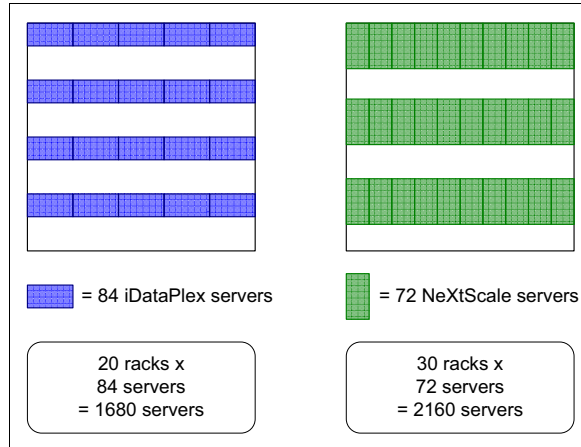


Figure 5-3 Increase in server density with NeXtScale compared to iDataPlex

The examples that are shown in Figure 5-6 are at the higher end of the density spectrum. They are meant to show what is possible with high-density designs.

5.4 Racks

In this section, we describe installing NeXtScale System in the 42U 1100mm Enterprise V2 Dynamic Rack because this rack is the rack that is used in our Intelligent Cluster solution. It is also the rack we recommend for NeXtScale System implementations. Examples of several best practices also are described. Other considerations for installing servers in other racks and other rack options also are described.

5.4.1 Rack Weight

The NeXtScale System can pose some weight challenges when deployed in raised floor environments. Table 5-8 shows the building block weights for the typical building block components that make up NeXtScale configurations.

Table 5-8 Building block weights for NeXtScale System components

Component	Weight
Chassis (Loaded)	112 kg
Network Switch	6.4 kg
Power distribution unit + power cord	11.7 kg
Cable weight per chassis (Ethernet + Power)	12 kg
Rack (Empty)	187 kg

Table 5-9 shows the total rack weights and floor loading for NeXtScale Systems. Typical configurations consist of each chassis having 12 compute nodes, and six 1300 W power supplies. Each rack consists of six NeXtScale n1200 enclosures, six network switches, and four PDUs.

Table 5-9 Total rack weights for NeXtScale System

Component	Weight
Shipping weight (Crated)	1056 kg
Weight (Uncrated)	1016 kg
Point load (four points per rack)	254 kg
Floor loading - Rack only (kg/m ²)	1540 kg/m ²
Floor loading - Rack + service area (kg/m ²)	605 kg/m ²

5.4.2 The 42U 1100mm Enterprise V2 Dynamic Rack

The 42U 1100mm Enterprise V2 Dynamic Rack is Lenovo’s leading open systems server rack. The word *dynamic* in the name means that it is shipped with equipment installed. As shown in Figure 5-7, it features removable outriggers to prevent it from tipping when it is moved with equipment that is installed in it.

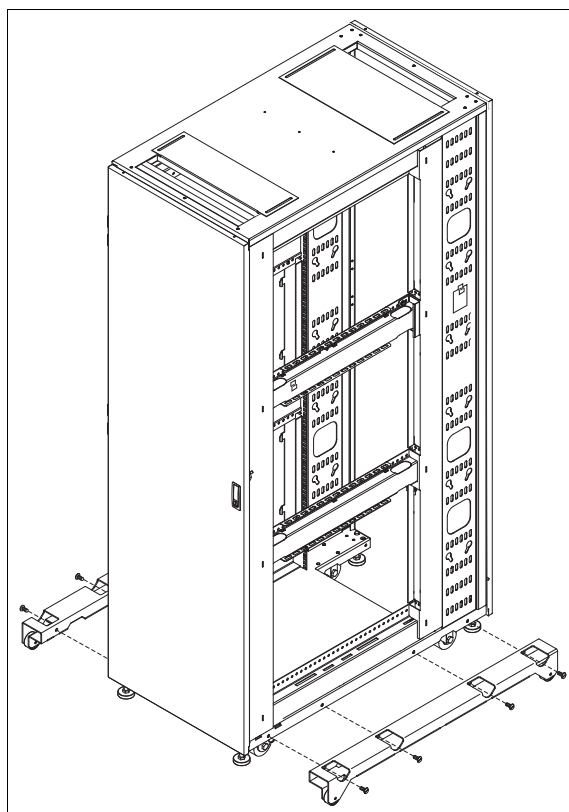


Figure 5-4 Outrigger removal or attachment

The rack features fully welded construction and is rated for a load of 953 kilograms (2100 pounds). The rack is one standard data center floor tile (600 mm) wide, and 1100 mm deep. When the optional Rear Door Heat Exchanger (which is described in 5.6, “Rear Door

Heat eXchanger” on page 89) is added, the rack is two standard data center floor tiles (1200 mm) deep.

The base model of the rack includes side panels. The expansion model of the rack includes the hardware that is used to join it to another rack, and no side panels. A row of racks (sometimes called a suite) can be created with one base model rack and one or more expansion racks; the side panels from the base rack are installed on each end of the row. Figure 5-8 shows joining two racks. Racks that are connected in this way match up with standard floor tile widths, which is not possible by moving the base model racks next to each other.

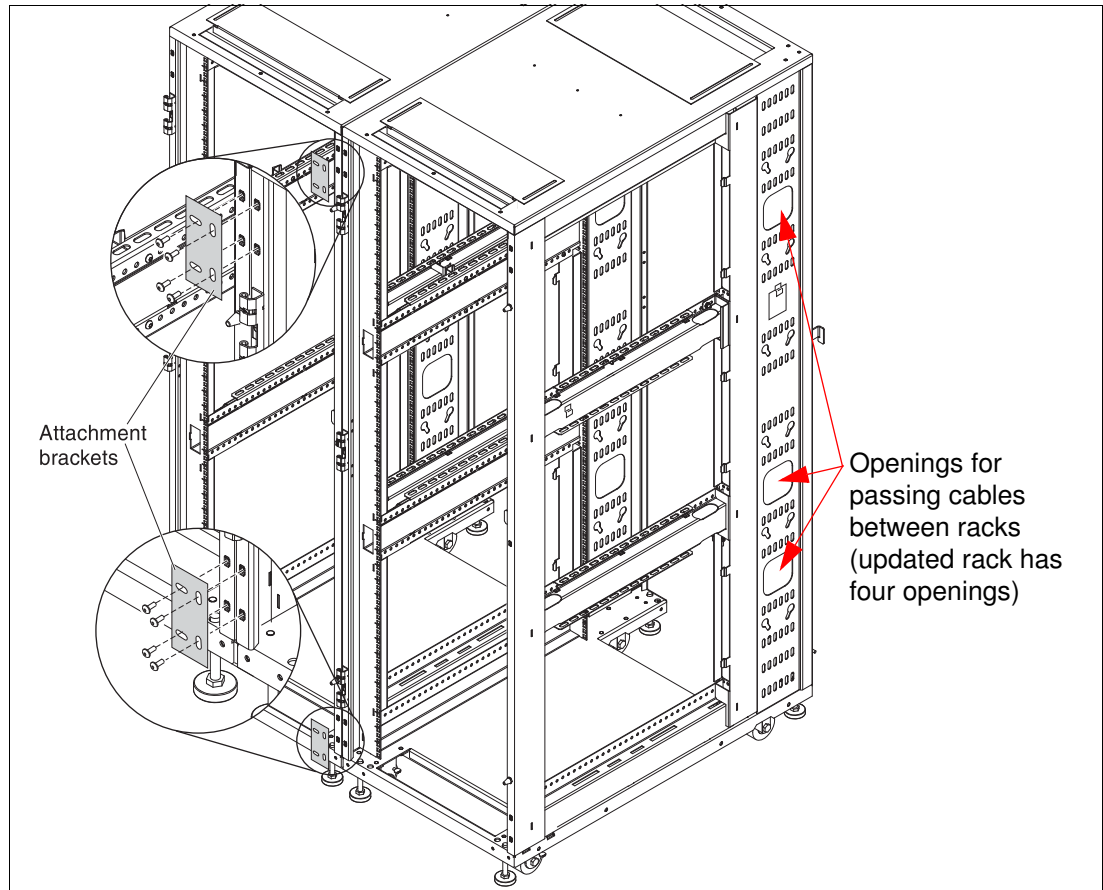


Figure 5-5 Joining two racks to make a row

The following features also are included:

- A front stabilizer bracket, which can be used to secure the rack to the floor, as shown in Figure 5-9.

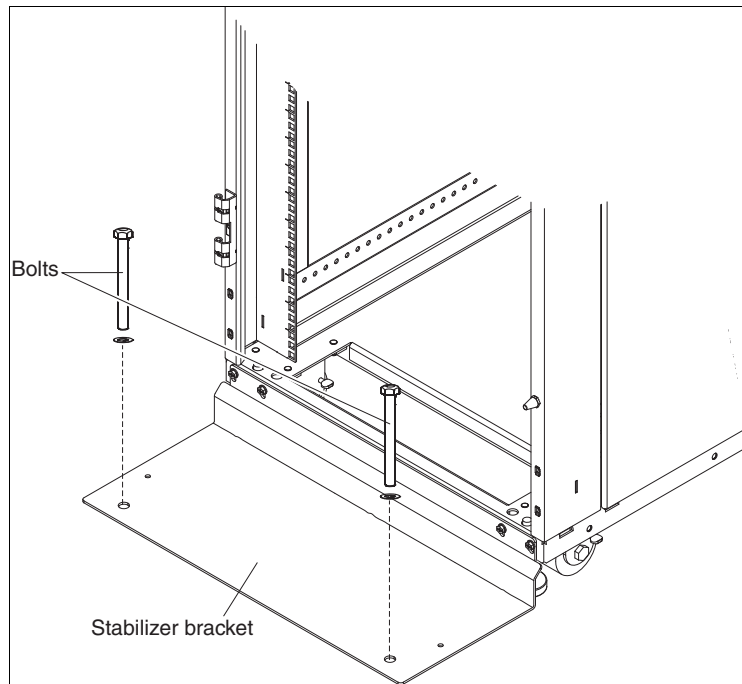


Figure 5-6 Rack with stabilizer bracket attached that is bolted to the floor

- As shown in Figure 5-10, a recirculation plate is used to prevent warm air that is entering from the rear of the rack from passing under the rack and into the front of the servers. This plate is not required if the stabilizer bracket is installed.

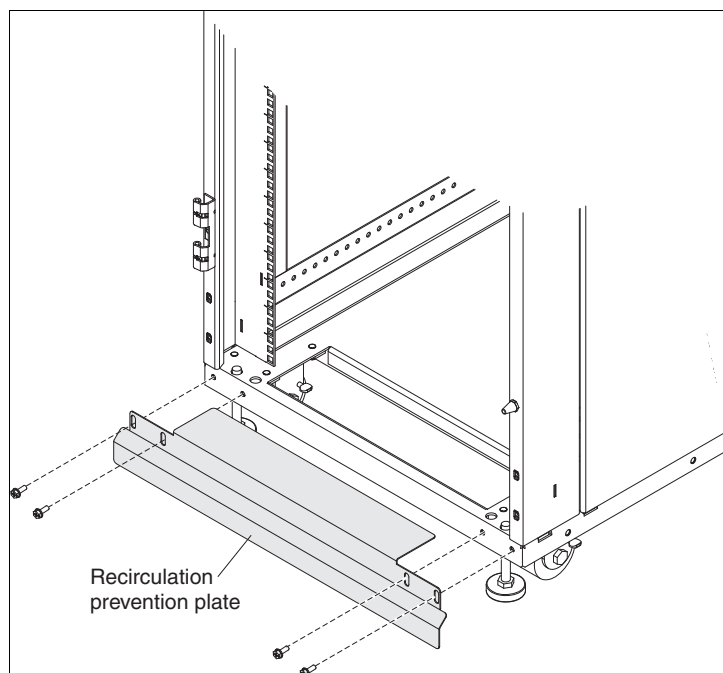


Figure 5-7 Attaching the recirculation prevention plate

Note: In addition to the recirculation plate or stabilizer bracket, another seal kit (part number 00Y3001) is required to prevent air recirculation through the opening at the front, bottom of the rack if a recessed switch is in the U space 1. For more information, see 5.4.5, “Rack options” on page 83.

- ▶ One-piece perforated front and rear doors with reversible hinges, so the doors can be made to open to the right or the left.
- ▶ Lockable front and rear doors, and side panels.
- ▶ Height of less than 80 inches, which enables it to fit through the doors of most elevators and doorways.
- ▶ Reusable, ship-loadable packaging. For more information about the transportation system, see this website:

<http://ibm.com/support/entry/portal/docdisplay?ln docid=migr-5091922>

- ▶ Six 1U vertical mounting brackets in the rear post flanges, which can be used for power distribution units, switches, or other 1U devices, as shown in Figure 5-11.

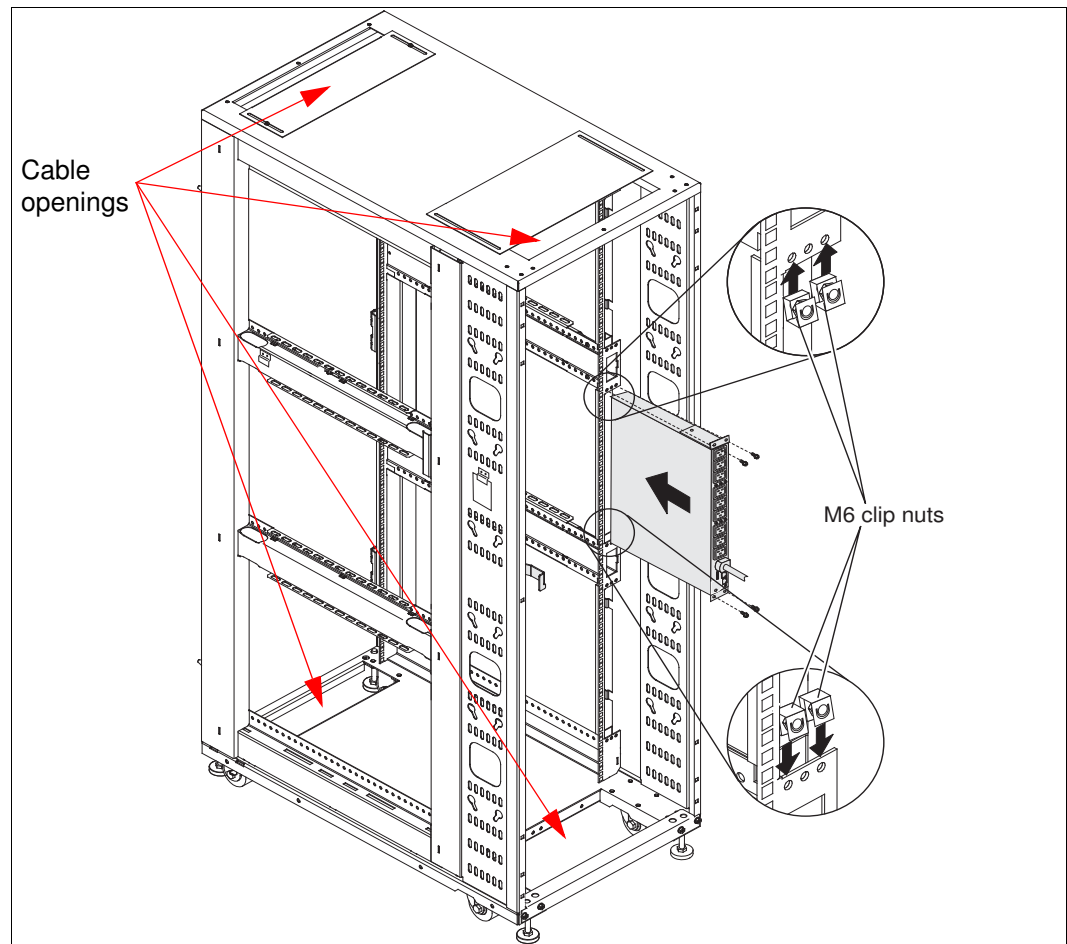


Figure 5-8 1U Power distribution unit mounted in 1U pocket on flange of rear post

- Two front-to-rear cable channels on each side of the rack, as shown in Figure 5-12.

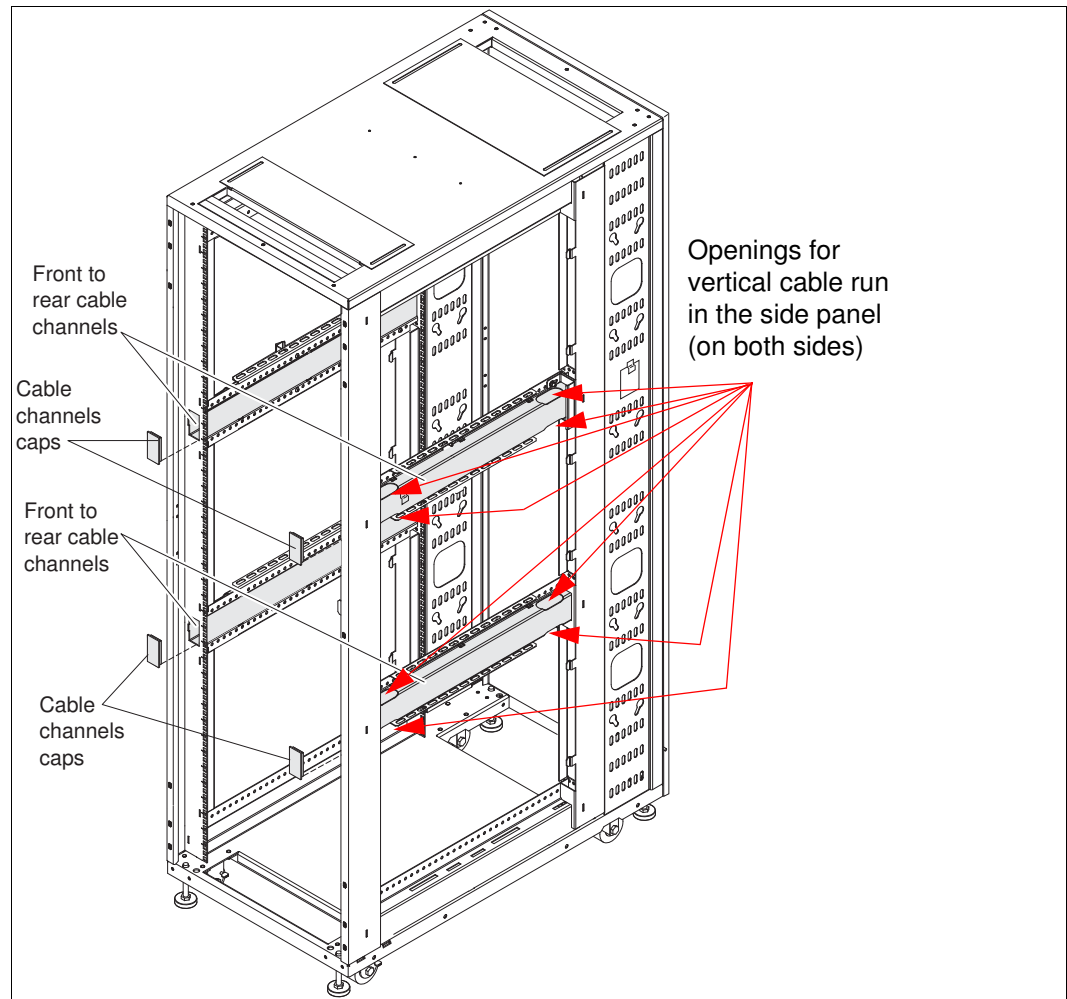


Figure 5-9 Front to rear cable channels

- Openings in the cable channels on each side of the rack to enable vertical cabling within the side panels, as indicated by arrows in Figure 5-12.
- Openings in the side walls behind the rear posts through which cables can be routed between racks in a row, as indicated by the arrows that are shown in Figure 5-8 on page 77. Also included are attachment points above and below these openings to hold cables out of the way and reduce cable clutter. New versions of this rack have four openings in each side wall.
- Front and rear cable access openings at the top and bottom of the rack, as indicated by the arrows that are shown in Figure 5-11 on page 79. The top cable access opening doors slide to restrict the size of the opening. For most effective use, do not tightly bundle cable that is passing through these openings. Instead, use the doors to flatten them in to a narrow row.
- Pre-drilled holes in the top of the rack for mounting third-party cable trays to the top of the rack.

The part numbers for the racks are listed in Table 5-12.

Table 5-10 Enterprise V2 Dynamic Rack part numbers

Part Number	Description
93634PX	42U 1100 mm Enterprise V2 Dynamic Rack
93634EX	42U 1100 mm Enterprise V2 Dynamic Expansions Rack

For more information about this rack, see the *Installation Guide*, which available at this website:

<http://www.ibm.com/support/entry/portal/docdisplay?lnocid=migr-5089535>

5.4.3 Installing NeXtScale System in other racks

The NeXtScale System chassis can be installed in other racks. The NeXtScale n1200 Enclosure can be installed in most industry-standard, four-post server racks. Figure 5-13 shows the dimensions of the NeXtScale n1200 Enclosure and included rail kit.

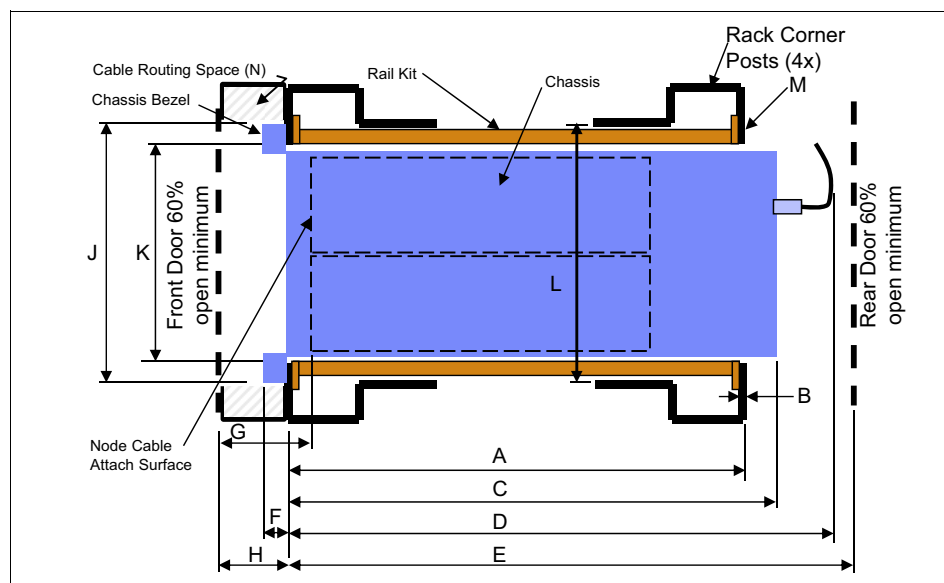


Figure 5-10 Dimensions for mounting the n1200 Enclosure rails and brackets (top view)

The following features are highlighted in Figure 5-13:

- The distance between the outside of front and outside of the rear rack EIA flanges can range 609 - 780 mm. This distance must be 719 mm to support the supplied rear shipping brackets when shipping is configured in a rack.
- The thickness of rack EIA flanges should be 2 mm - 4.65 mm. The supplied cage and clip nuts do not fit on material thicker than 3 mm.
- The distance from front EIA flange to the rear of the system is 915 mm (including handles and latches, the distance is approximately 935 mm).
- The distance from front EIA flange to the bend radius of the rear cables is 980 mm.
- The minimum distance from the front EIA flange to the closest features on the inside of the rear door is 985 mm.
- The distance from front EIA flange to the front bezel of the system is 35 mm.

- G. A recommended minimum distance from front of node cable plug surface to inside of front door is 120 mm.
- H. A recommended minimum distance from front EIA flange to the closest features on the inside of the front door is approximately 80 mm.
- J. The width of the front of the system bezel is 482 mm.
- K. The minimum horizontal opening of the inside of the rack at the front and rear EIA flanges is 450 mm.
- L. The minimum width between the required internal structure to the rack to mount rail kits is 475 mm (481 mm within 50 mm of the front and rear EIA flanges).
- M. Mounts use 7.1 mm round or 9.5 mm square hole racks. Tapped hole racks with standard rail kit are not supported.
- N. NeXtScale System data and management cables attach to the front of the node (chassis). Cable routing space must be provided at the front of the rack, inside or outside the rack. Cable space (N) should from the bottom to the top of the rack and be approximately 25 x 75 mm in size. Some cable configurations might require more cable routing space.

We suggest installing the chassis 1 or 2 rack units from the bottom to allow space for cables and the power line cords to exit the rack without blocking service access to the chassis.

5.4.4 Shipping the chassis

Lenovo supports shipping the NeXtScale n1200 Enclosure in an 42U 1100mm Enterprise V2 Dynamic Rack if the shipping brackets that are supplied with the chassis are installed. Shipping the chassis in any other rack is at your discretion.

When it is installed in a rack and moved or shipped from one location to another, the NeXtScale System shipping brackets must be reattached. For more information, see the documentation that is included with the chassis for the installation of the shipping brackets.

Cooling considerations

It is important that the NeXtScale n1200 Enclosure be installed in an environment that results in proper cooling. The following rack installation considerations are important:

- ▶ The rack must have front (server node side) to back airflow.
- ▶ The rack must have front and rear doors with at least 60% open area for airflow.
- ▶ To allow adequate airflow into and out of the chassis, adhere to the recommended minimum distances to the front and rear doors, as shown in Figure 5-13 on page 81.
- ▶ It is important to prevent hot air recirculation from the back of the rack to the front. The following points should be considered:
 - All U spaces must be occupied at the front by a device or a blank filler panel.
 - All other openings should be covered, including air openings around the EIA flanges (rack posts) and cable passage ways.
 - If multiple racks are arranged in a row, gaps between the rack front EIA flanges must be blocked to prevent hot air recirculation.
- ▶ Seal openings under the front of the racks.

5.4.5 Rack options

The rack options that are listed in Table 5-13 are available for the Enterprise V2 Dynamic Racks and other racks.

Table 5-11 Rack option part numbers

Part number	Description
Monitor kits and keyboard trays	
17238BX	1U 18.5-inch Standard Console
17238EX	1U 18.5-inch Enhanced Media Console
172317X	1U 17-inch Flat Panel Console Kit
172319X	1U 19-inch Flat Panel Console Kit
Console switches	
1754D2X	Global 4x2x32 Console Manager (GCM32)
1754D1X	Global 2x2x16 Console Manager (GCM16)
1754A2X	Local 2x16 Console Manager (LCM16)
1754A1X	Local 1x8 Console Manager (LCM8)
Console cables	
43V6147	Single Cable USB Conversion Option (UCO)
39M2895	USB Conversion Option (four Pack UCO)
39M2897	Long KVM Conversion Option (four Pack Long KCO)
46M5383	Virtual Media Conversion Option Gen2 (VCO2)
46M5382	Serial Conversion Option (SCO)

The options that are listed in Table 5-14 are unique to configuring NeXtScale environments.

Table 5-12 Racking parts that are specific to NeXtScale

Part number	Description
00Y3011	1U Pass Through Bracket
00Y3016	Front cable management bracket kit
00Y3026	Cable Routing Tool
00Y3001	Rack and Switch Seal Kit

The 1U Pass Through Bracket is shown in Figure 5-14. The front for the component has brushes to block the airflow around any cables that are passed through it. It can also serve to block air flow around a switch that is recessed in the rack (for cable routing reasons), that pass around a blank filler panel.

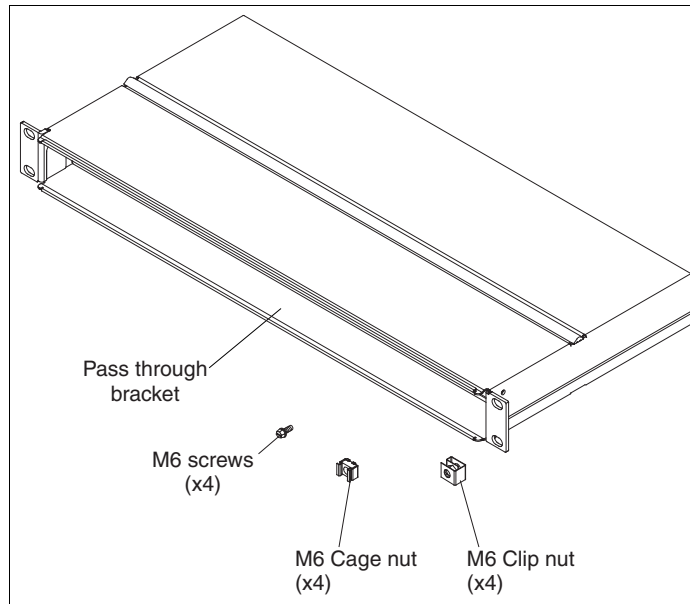


Figure 5-11 1U filler with brushes, allows cables through but blocks air flow

The Front Cable Management Bracket kit (part number 00Y3016) that attaches to the front of the rack is shown in Figure 5-15. This kit includes four brackets, which are enough for one rack (two brackets are installed on each side of the rack).

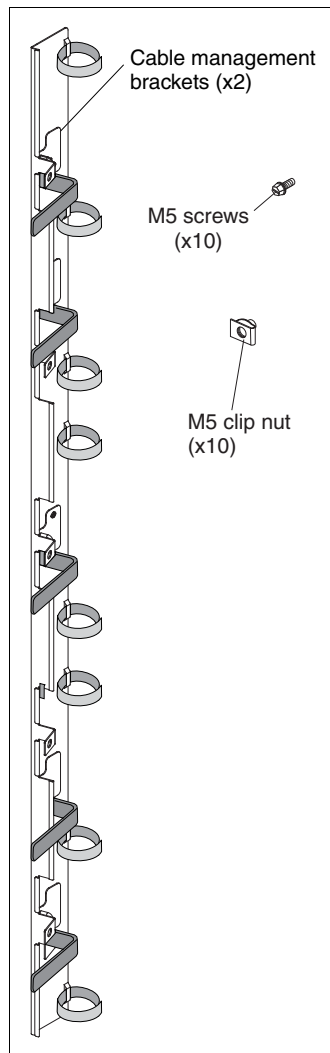


Figure 5-12 Cable management bracket

The Cable Routing Tool (part number 00Y3026) that is shown in Figure 5-15 is a plastic rod to which a cable or cables can be attached by using a hook-and-loop fastener. After it is assembled, the tool is used to pull the cables through the cable channels.

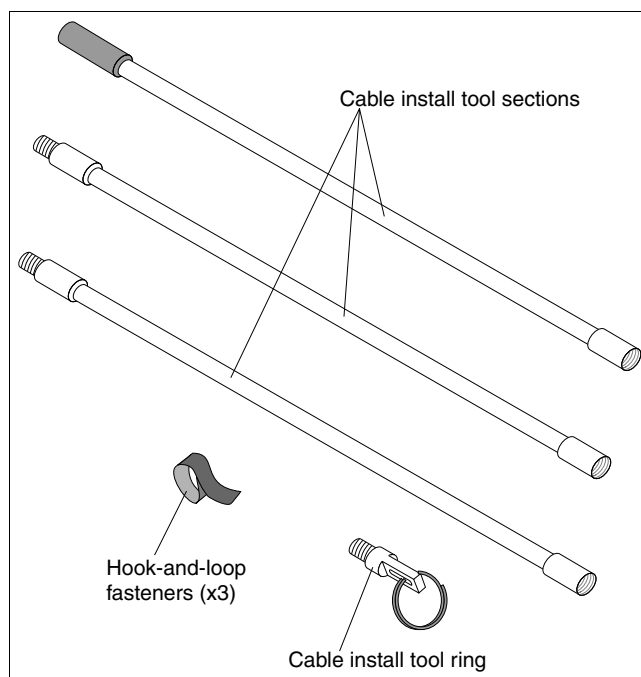


Figure 5-13 Cable routing tool

The Rack and Switch Seal Kit (part number 00Y3001) has the following purposes:

- Provides a means to seal the opening in the bottom front of the 42U 1100 mm Enterprise V2 Dynamic Rack. The opening at the bottom front of the rack must be sealed if a switch is at the front of the rack in U space one.
- Provides air sealing of switch mounting rails of switches that are mounted at the front of a rack and are recessed behind the rack mounting rails. The seal kit includes enough switch seals for six switches.

Figure 5-17 shows the components of the Rack and Switch Seal Kit.

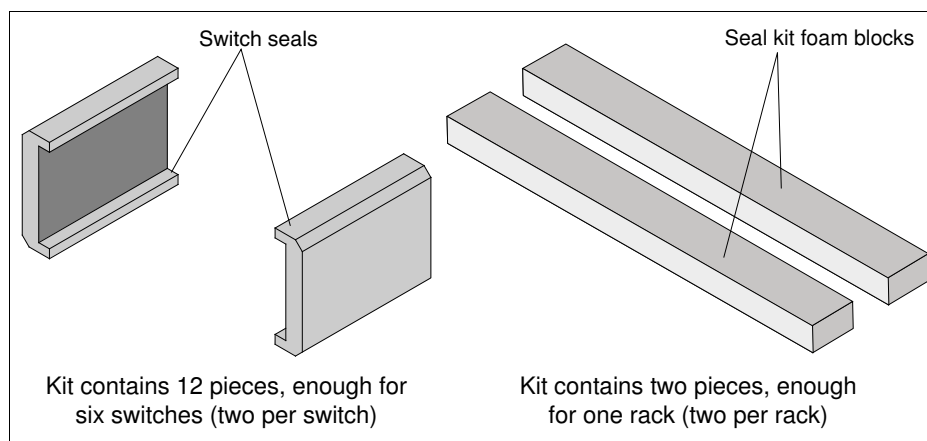


Figure 5-14 Rack and Switch Seal Kit contents

Figure 5-18 shows where these pieces are used.

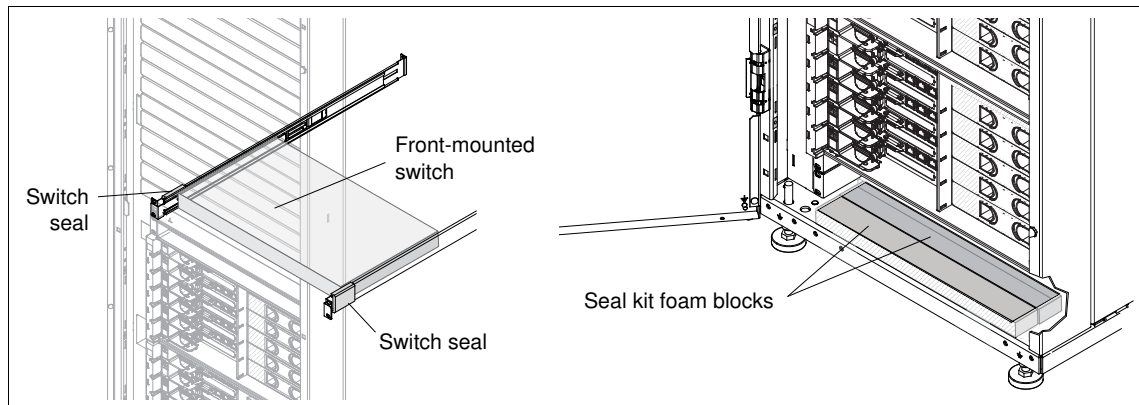


Figure 5-15 Placement of the components of the Rack and Switch Seal Kit

5.5 Cable management

NeXtScale System was designed with serviceability in mind. All of the compute node connectivity is at the front so nodes can be serviced without any guesswork as to which cables belong to which nodes. Managing and routing all of this cabling can cause problems; however, Lenovo developed NeXtScale System with options to make cable management simple.

Cable routing

As seen in Figure 5-12 on page 80, the 42U 100 mm Enterprise V2 Dynamic Rack contains two front to rear cable channels on each side of the rack so routing cables to the back of the rack does not waste any U space in the rack. Having these multiple openings on each side of the rack at different heights reduces the overall size of the cable bundles that are supporting the compute nodes.

Cables can be routed between racks in a row through openings in the side walls behind the rear posts, as indicated by arrows in Figure 5-8 on page 77. Also included are attachment points above and below these openings to hold cables out of the way and reduce cable clutter. New versions of this rack have four openings in each side wall.

Front and rear cable access openings are available at the top and bottom of the rack, as indicated by arrows in Figure 5-11 on page 79. The top cable access opening doors slide to restrict the size of the opening. For most effective use, do not tightly bundle cable that is passing through these openings. Instead, use the doors to flatten them in to a narrow row.

Pre-drilled holes are available in the top of the rack for mounting third-party cable trays to the top of the rack.

Cable management brackets

An optional cable management bracket kit is available to help with cable management of the cable bundles at the front of the rack. This kit consists of six brackets (enough for three chassis) that attach to the front left and right of the chassis. To use the bracket, the front door of the rack often must be removed because the bracket can extend beyond the front of the rack. Figure 5-19 shows the cable bracket.

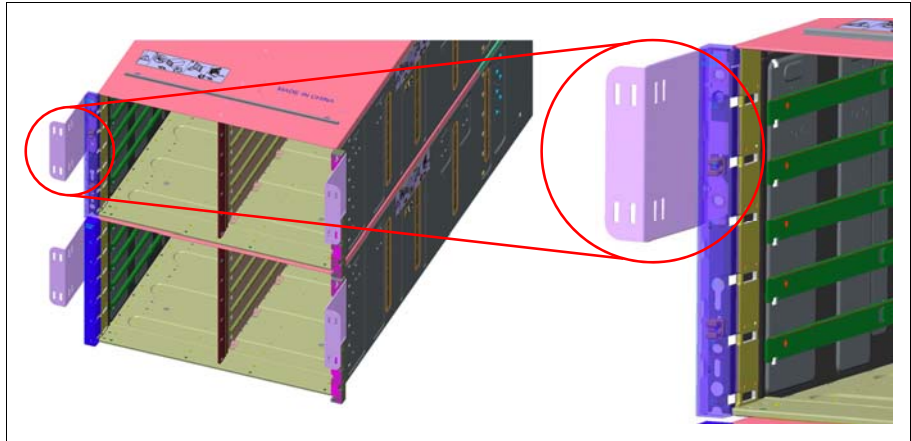


Figure 5-16 Cable management bracket kit for third-party racks, part 00Y3040

The part number of the Cable Management Bracket kit is listed in Table 5-15.

Table 5-13 Cable Management Bracket Kit part number

Part number	Description
00Y3040	Cable management bracket kit (contains 6 brackets and 20 hook-and-loop fasteners)

Figure 5-20 shows a top view of the bracket and possible location for cable bundles.

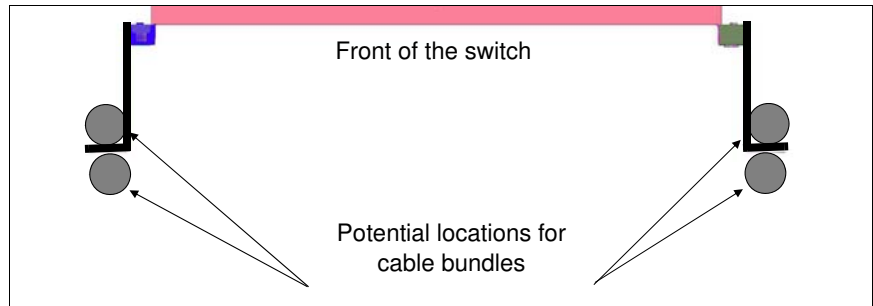


Figure 5-17 Top view of cable management bracket kit, showing cable bundles

Another option for managing cable bundles at the front of the rack is the Front Cable Management Bracket kit (part number 00Y3016). This bracket assembly attaches to the front of the rack, as shown in Figure 5-15 on page 85. This kit includes four brackets, which are enough for one rack (two brackets are installed on each side of the rack).

Cable routing support

The Cable Routing Tool (part number 00Y3026) that is shown in Figure 5-16 on page 86 is a plastic rod to which a cable or cables can be attached by using a hook-and-loop fastener. After it is assembled, the tool is used to pull the cables through the cable channels. This process helps make changes to the rack cabling much easier and reduces the potential for damage to the cabling by not forcing cables through confined environments.

5.6 Rear Door Heat eXchanger

The heat exchanger is a water-cooled door that is mounted on the rear of an 42U 1100 mm Deep Dynamic Rack Type 9363 to cool the air that is heated and exhausted by devices inside the rack. A supply hose delivers chilled, conditioned water to the heat exchanger. A return hose delivers warmed water back to the water pump or chiller. In this document, this configuration is referred to as a *secondary cooling loop*. The primary cooling loop supplies the building chilled water to secondary cooling loops and air conditioning units. The rack on which you install the heat exchanger can be on a raised floor or a non-raised floor. Each heat exchanger can remove 100,000 BTU per hour (or approximately 30,000 watts) of heat from your data center.

The part number for the Rear Door Heat eXchanger for 42U 1100 mm rack is shown in Table 5-16.

Table 5-14 Part number for Rear Door Heat eXchanger for 42U 1100 mm rack

Part number	Description
175642X	Rear Door Heat eXchanger for 42U 1100 mm Enterprise V2 Dynamic Racks

For more information, see *Rear Door Heat eXchanger V2 Type 1756 Installation and Maintenance Guide*, which is available at this website:

<http://www.ibm.com/support/entry/portal/docdisplay?ln docid=migr-5089575>

Rear Door Heat eXchanger V2 overview

Table 5-17 lists the specifications of the Rear Door Heat eXchanger V2 type 1756.

Table 5-15 Rear Door Heat eXchanger V2 specifications

Parameter	Specification
Door dimensions	
Depth	129 mm (5.0 in.)
Height	1950 mm (76.8 in.)
Width	600 mm (23.6 in.)
Door weight	
Empty	39 kg (85 lb)
Filled	48 kg (105 lb)

Parameter	Specification
Pressure	
Normal operation	<137.93 kPa (20 psi)
Maximum	689.66 kPa (100 psi)
Volume	
Water volume	9 liters (2.4 gallons)
Flow rate	
Nominal flow	22.7 lpm (6 gpm)
Maximum flow	56.8 lpm (15 gpm)
Water Temperature	
Minimum (non-condensing)	Above dew point
ASHRAE Class 1	18 °C +/- 1 °C (64.4 °F +/- 1.8 °F)
ASHRAE Class 2	22 °C +/- 1 °C (71.6 °F +/- 1.8 °F)

The Rear Door Heat eXchanger (RDHX) also includes the following features:

- ▶ Attaches in place of the perforated rear door and adds 100 mm, which makes the overall package 1200 mm (the depth of two standard data center floor tiles).
- ▶ The doors use 3/4-inch quick connect couplers, which include automatic valves that restrict water leakage (often a few drops at most) when the doors are connected or disconnected.
- ▶ Each door has a capacity of 9 liters (2.4 US gallons), and supports flow rates of 22.7 liters (6 US gallons) to 56.8 liters (15 US gallons) per minute.
- ▶ The doors have no moving parts; the fans in the equipment move air through the heat exchanger as easily as a standard rack door.
- ▶ If the water flow is disrupted, the rack reverts to standard air cooling.

Rear Door Heat eXchanger performance

Each door can remove 100% of the heat that is generated by servers that use 30 kW of power and 90% of the heat that is generated by servers that use 40 kW. It removes the heat by using 18 °C (64 °F) water at a 27 °C (81 °F) server inlet air temperature.

Figure 5-21 shows more information about the capability of this rear door heat exchanger.

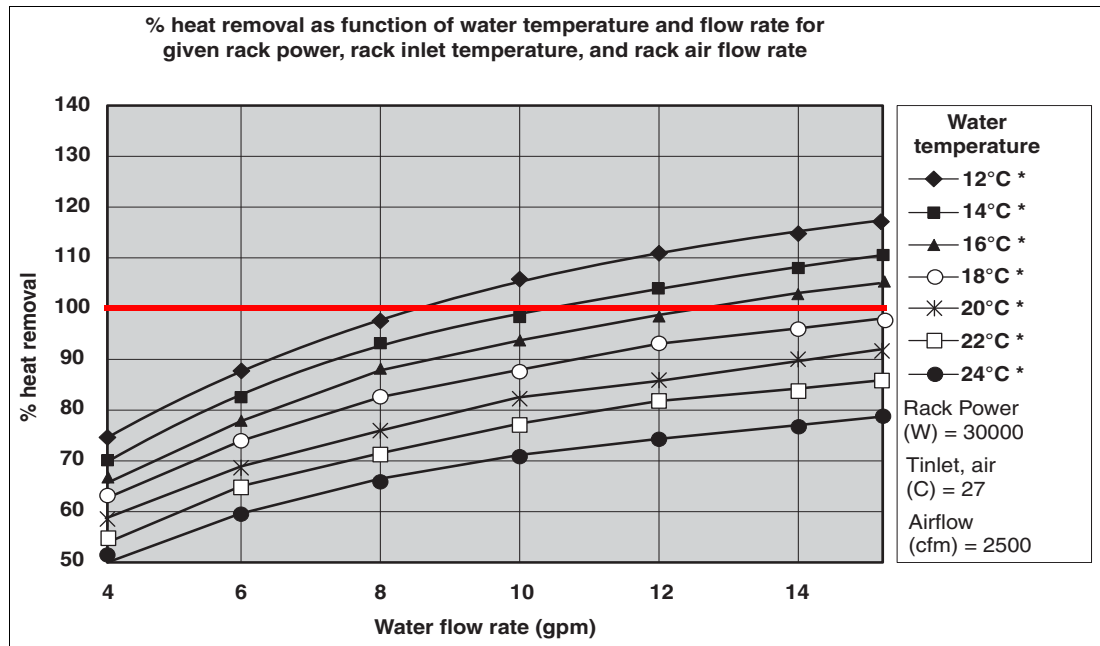


Figure 5-18 Heat removal performance with a 30 kW load

Although even more heat can be extracted if the water is cooler, the water temperature cannot be below the dew point in the server room or condensation forms on the rear door.

Some standard computer room air conditioning often is provisioned to control humidity and enable doors to be disconnected for maintenance or other requirements

The reduced air conditioning requirement typically saves about 1 KW per rack that is used to compress refrigerant and move air.

The reduction in air conditioner noise, which is coupled with the acoustic dampening effect of the heat exchangers and the decrease in high velocity cold air, makes the data center environment less hostile.

5.7 Top-of-rack switches

NeXtScale System does not include integrated switching in the chassis, unlike BladeCenter or Flex System. Integrated switching was not included to maximize the modularity and flexibility of the system and prevent any chassis-level networking contention. The Intelligent Cluster offering includes various switches from Lenovo Networking and others that can be configured and ordered as an integrated solution. Users who are building solutions (including NeXtScale System) can combine them with switches of their own choosing. In the following sections, we describe the switches that are available from Lenovo.

Note: There is no rule that smaller rack mountable switches (typically 1U or 2U) must be mounted in the top of a rack, but “top-of-rack” is the common name they were given. This idea is in contrast to larger switches that are deployed for rows of rack-mounted servers, which are referred to as “end-of-row” switches; or the often large, modular switches, which also often contain routing functionality, which are commonly called “core switches”.

5.7.1 Ethernet switches

NeXtScale System features forward-facing cabling. To connect cables from the NeXtScale servers to Ethernet switches, it is easiest to mount the switches in the racks with the switch ports facing the front of the racks as well.

It is important that the cooling air for the switches flow from the front (port side) of the switch to the rear (non-port side). The switches that are listed in Table 5-18 on page 94 meet this criteria. Switches that are cooled from rear to front can be used, but these switches must be mounted facing the rear of the rack. Also, all the cables from the servers are routed from the front of the rack to the back to connect.

Table 5-16 Top-of-rack switches

Part number	Description
1 Gb top-of-rack switches	
715952F	Lenovo RackSwitch™ G8052 (Front to Rear)
10 Gb top-of-rack switches	
7159BF7	Lenovo RackSwitch G8124E (Front to Rear)
715964F	Lenovo RackSwitch G8264 (Front to Rear)
7159DFX	Lenovo RackSwitch G8264CS (Front to Rear)
7159CFV	Lenovo RackSwitch G8272 (Front to Rear)
7159GR5	Lenovo RackSwitch G8296 (Front to Rear)
40 Gb top-of-rack switches	
7159BFX	Lenovo RackSwitch G8332 (Front to Rear)
Rail kit	
00CG089	Recessed 19-inch 4-Post Rail Kit

The Recessed 19-inch, 4-Post Rail Kit (00CG089) locates the front of the switch 75 mm behind the front posts of the rack to provide more cable bend radius. The rail kit is recommended for mounting 1U switches in racks with NeXtScale chassis.

5.7.2 InfiniBand switches

The InfiniBand switches (as listed in Table 5-19) are available as part of the Intelligent Cluster offering. As with the Ethernet switches that require cooling from the ports to the back of the switch, the InfiniBand switches need the same air flow path. However, on the InfiniBand switches, it is referred to as opposite port side exhaust (oPSE). It is recommend that the System Networking Recessed 19-inch 4-Post Rail Kit is installed for cable bend radius reasons.

Table 5-17 InfiniBand switch feature codes

Feature code	Description
A2EZ	Mellanox SX6036 FDR10 InfiniBand Switch (oPSE)
A2Y7	Mellanox SX6036 FDR14 InfiniBand Switch (oPSE)
6676	Intel 12200 QDR IB Redundant Power Switch (oPSE)

Feature code	Description
6925	Intel 12200 QDR InfiniBand Switch (oPSE)

5.7.3 Fibre Channel switches

For I/O intensive applications, 8 Gb and 16 Gb Fibre Channel networks (which are compatible with 8 Gb and 4 Gb storage devices) are popular because of their high I/Os per second (IOPS) capability and high reliability. In larger clusters or systems with high I/O demands, there are several nodes that are connected to SAN storage, which then share storage via a parallel file system (such as IBM GPFS) to the rest of the servers. Table 5-20 on page 95 lists the current 16 Gb Fibre Channel switch offerings. These switches should be mounted facing the rear of the rack because they all have rear to front (port side) cooling air flow. The fiber optic connections must pass from the adapters that are installed in the front of the NeXtScale nodes to the ports at the rear of the rack.

Table 5-18 Fibre Channel switch part numbers

Part number	Description
16 Gb Fibre Channel switches	
2498F24	IBM System Storage SAN24B-5
2498F48	IBM System Storage SAN48B-5
2498F96	IBM System Storage SAN96B-5
8 Gb Fibre Channel switches	
249824E	IBM System Storage SAN24B-4 Express
241724C	Cisco MDS 9124 Express
2417C48	Cisco MDS 9148 for IBM System Storage
2498B80	IBM System Storage SAN80B-4

For more information about Fibre Channel switches, see this website:

<http://www.ibm.com/systems/networking/switches/san/>

5.8 Rack-level networking: Sample configurations

In this section, we describe the following sample configurations that use NeXtScale systems:

- ▶ InfiniBand non-blocking
- ▶ InfiniBand 50% blocking
- ▶ 10 Gb Ethernet configuration with one port per node
- ▶ 10 Gb Ethernet configuration with two ports per node

The location of the chassis and the switches within the rack are shown in a way that optimizes the cabling of the solution. The chassis and switches are color-coded to indicate which InfiniBand or Ethernet switches support which chassis.

Management network: Management networking is the same for all configurations. For more information, see 5.8.5, “Management network” on page 100.

For more information about networking with NeXtScale System, see *NeXtScale System Network and Management Cable Guide*, which is available at this website:

<http://ibm.com/support/entry/portal/docdisplay?lnv0-POWINF>

5.8.1 Non-blocking InfiniBand

Figure 5-24 shows a non-blocking InfiniBand configuration. On the left side is a rack with six chassis for a total of 72 compute nodes. Four 36-port InfiniBand switches and two 48-port 1 Gb Ethernet switches are used.

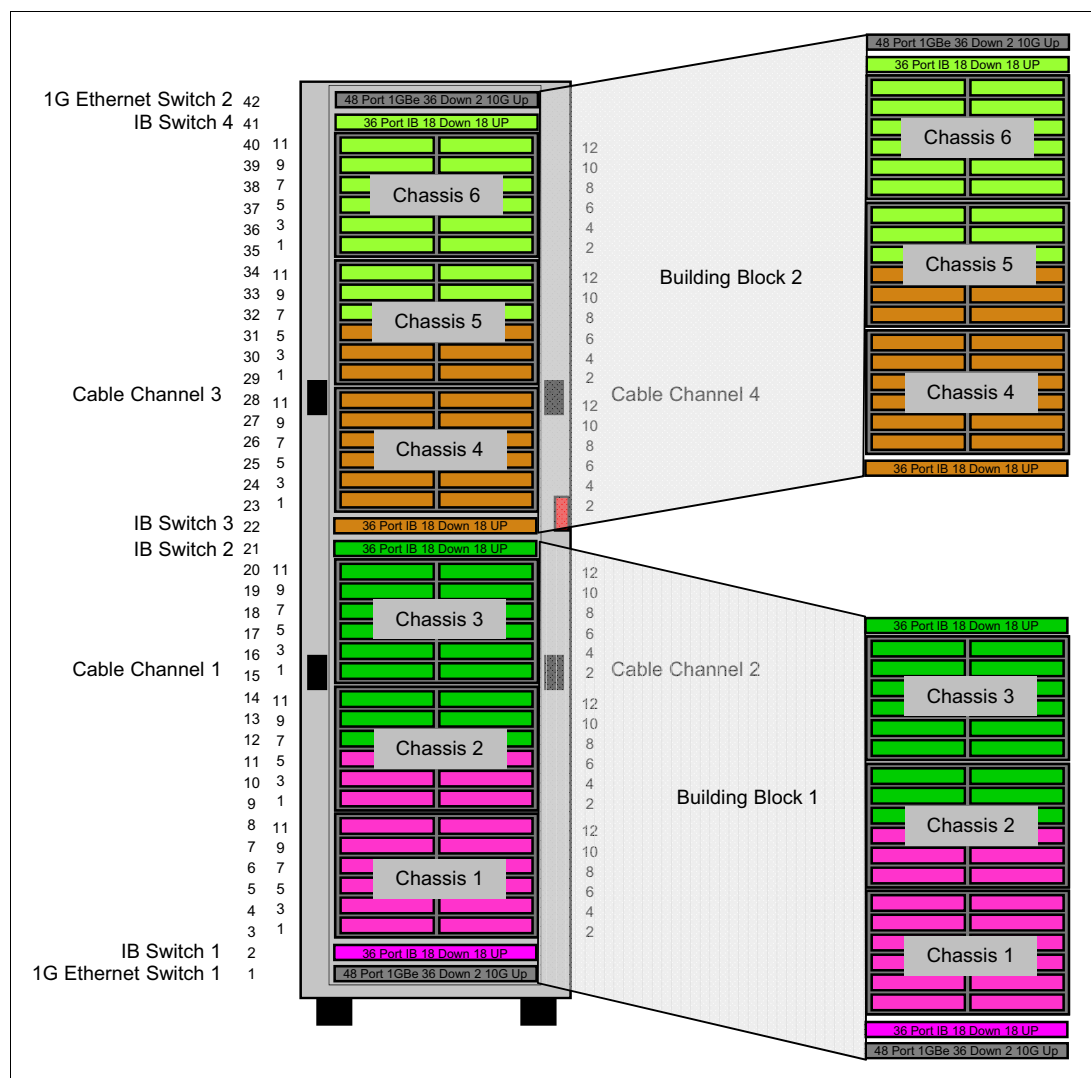


Figure 5-19 Non-blocking InfiniBand

5.8.2 A 50% blocking InfiniBand

Figure 5-25 shows a 50% blocking (two node ports for every uplink port) InfiniBand configuration. On the left side is a rack with six chassis for a total of 72 CPU nodes. Three 36-port InfiniBand switches and two 48-port 1 Gbps Ethernet switches provide the network connectivity.

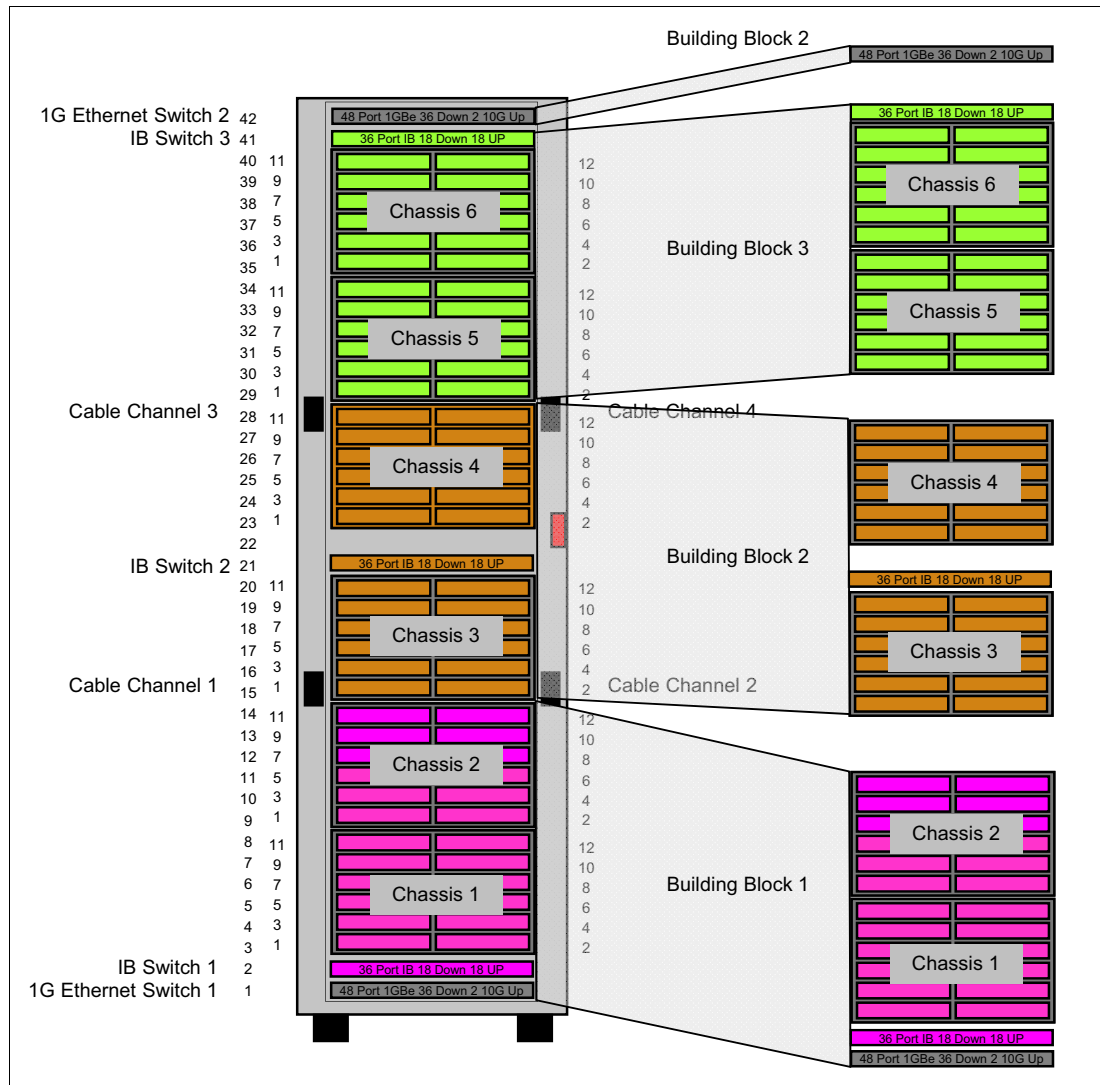


Figure 5-20 A 50% Blocking InfiniBand

Filler panel: Filler panels (part number 00Y3011) are placed in rack unit 21 and 41 to prevent hot air recirculation.

5.8.3 10 Gb Ethernet, one port per node

Figure 5-26 shows a network with one 10 Gb Ethernet connection per compute node. On the left side is a rack with six chassis for a total of 72 CPU nodes. Two 48 port 10 Gb switches and two 48 port 1 Gbps Ethernet switches provide the network connectivity.

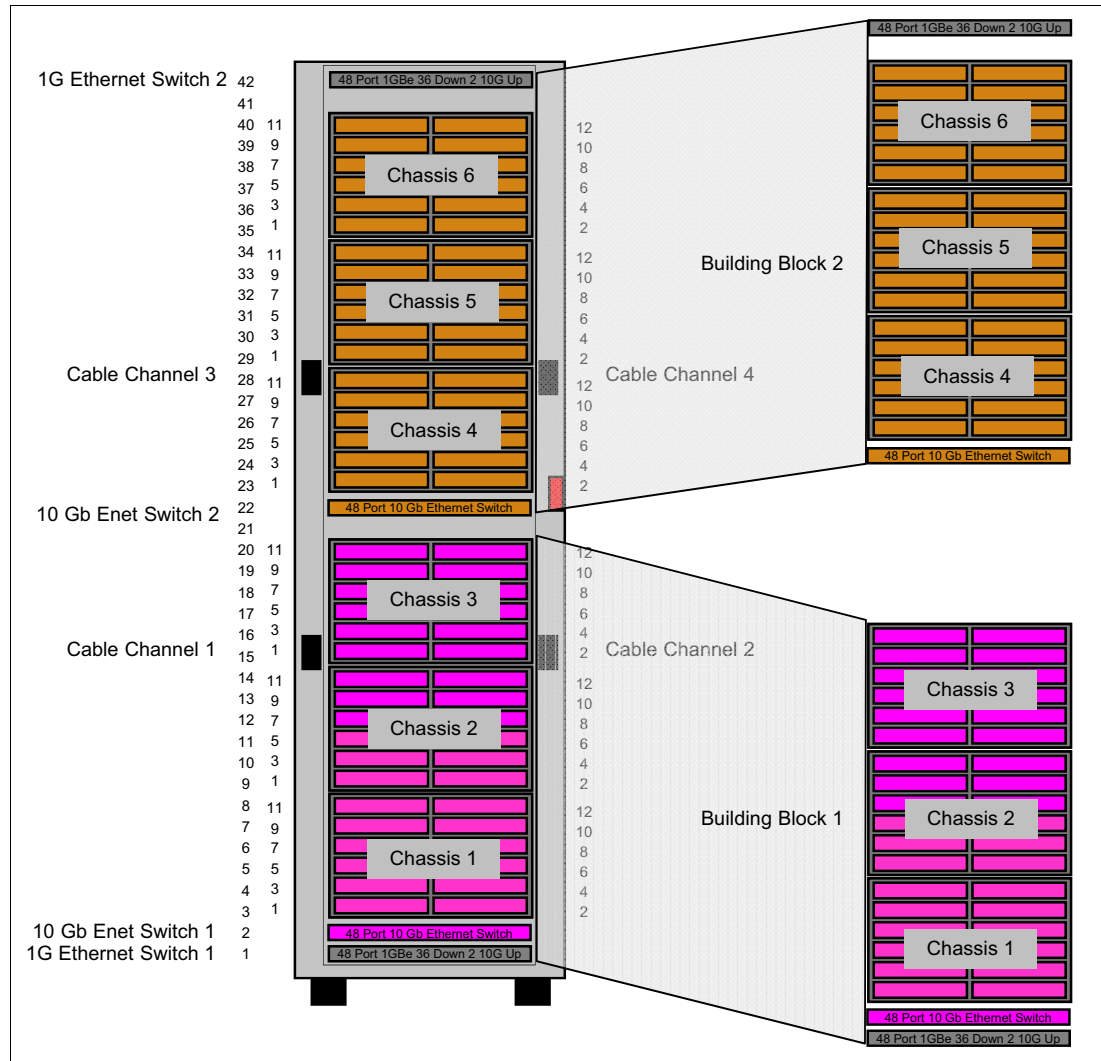


Figure 5-21 10 Gb Ethernet, one port per node configuration

Filler panel: 1U filler (part number 00Y3011) is placed in rack units 21 and 41 to prevent hot air recirculation.

5.8.4 10 Gb Ethernet, two ports per node

Figure 5-27 shows a network that consists of two 10 Gb Ethernet connections per compute node. On the left side is a rack with six chassis for a total of 72 compute nodes. Three 48-port 10 Gb Ethernet switches and two 48-port 1 Gb Ethernet switches provide the network connectivity.

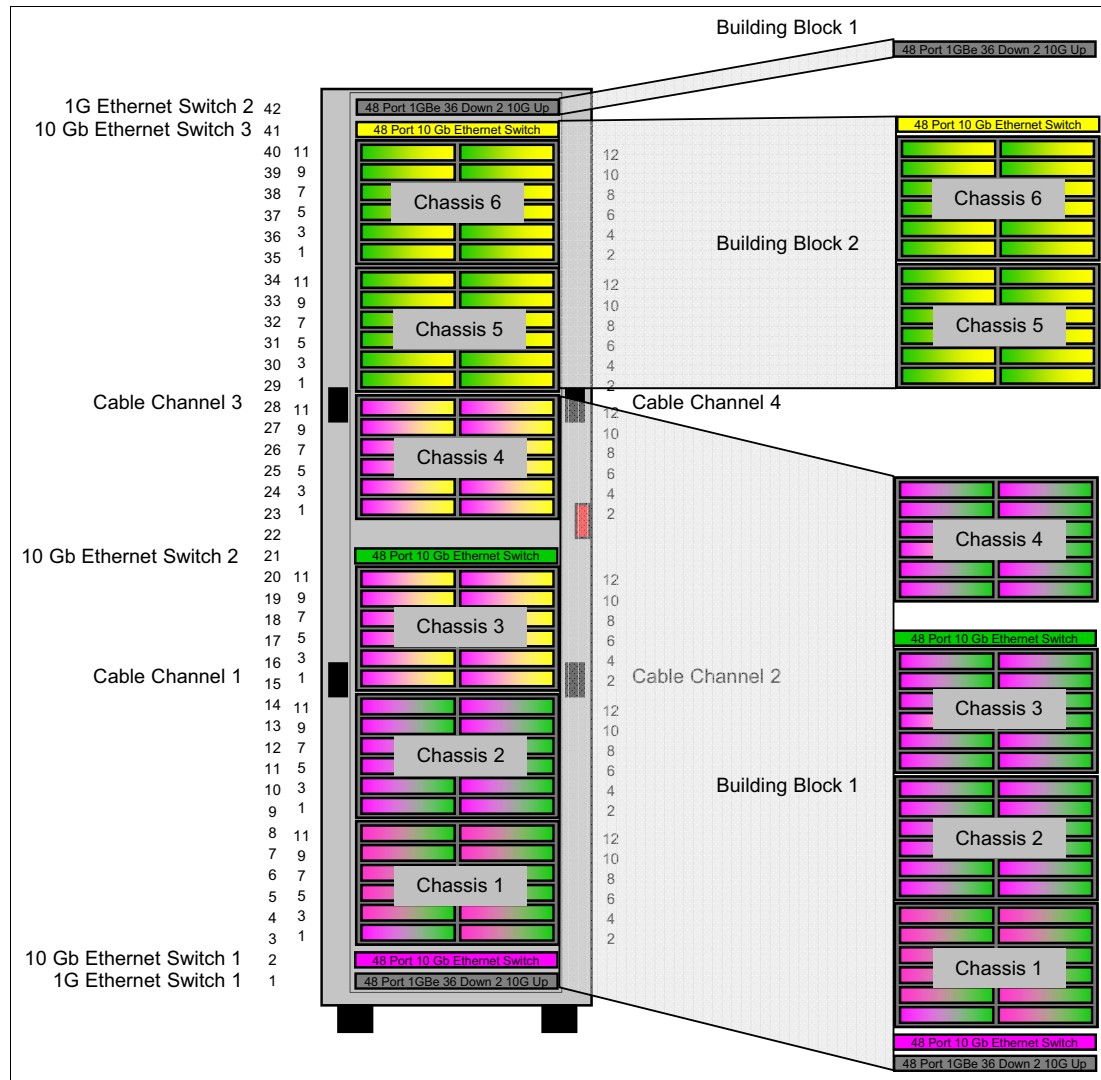


Figure 5-22 10 Gb Ethernet, two ports per compute node

The location of the chassis and the switches within the rack are shown in a way that optimizes the cabling of the solution. The chassis and switches are color-coded to indicate which InfiniBand or Ethernet switches support which chassis.

In Figure 5-27, each node has two colors, which indicates that each node is connected to two different switches to provide redundancy.

Filler panel: A 1U filler (part number 00Y3011) is placed in rack unit 22 to prevent hot air recirculation.

5.8.5 Management network

Figure 5-28 shows the 1 Gb Ethernet management network for a solution with four monitored PDUs. The 1 Gb management switches are shown in rack units 1 and 42.

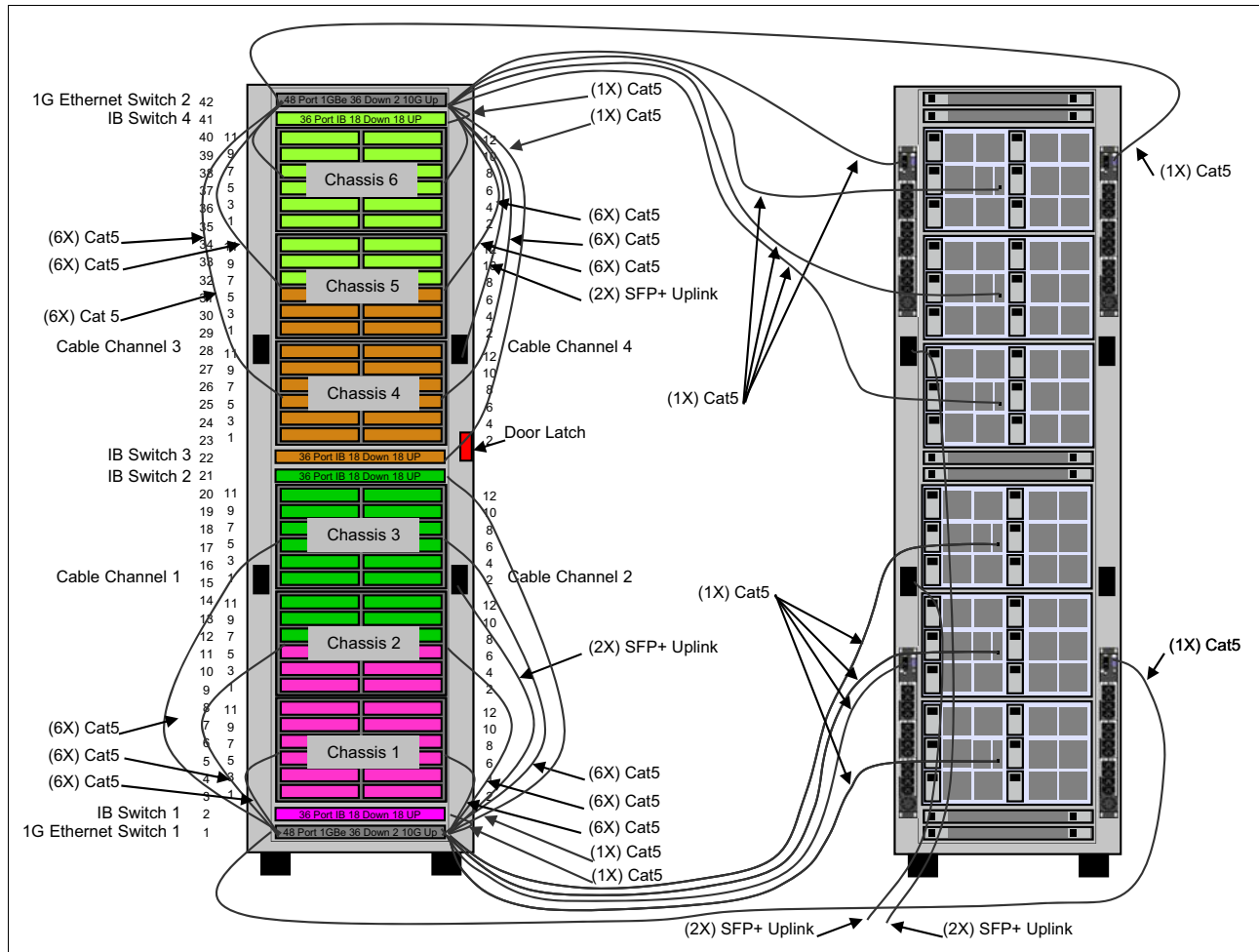


Figure 5-23 Management network

Each chassis has one management port that plugs to the Fan and Power Controller, which is at the rear of the chassis. Each PDU can also have a management port.

Note: The management cables that connect to devices at the rear of the chassis should be routed to the front of the chassis via the cable channels.

Factory integration and testing

NeXtScale System can be fulfilled through regular channels or as part of a fully integrated solution with Lenovo Intelligent Cluster. Lenovo provides factory integration and testing as part of the Intelligent Cluster offering.

This chapter describes Lenovo Intelligent Cluster, what is provided by Lenovo factory integration, the testing that is performed, and the documentation that is supplied.

This chapter includes the following topics:

- ▶ 6.1, “Lenovo Intelligent Cluster” on page 102
- ▶ 6.2, “Lenovo factory integration standards” on page 102
- ▶ 6.3, “Factory testing” on page 103
- ▶ 6.4, “Documentation provided” on page 105

6.1 Lenovo Intelligent Cluster

Deploying solutions for Technical Computing, High Performance Computing (HPC), Analytics, and Cloud environments can place a significant burden on IT staff. Through Intelligent Cluster, Lenovo brings its expertise in HPC design, deployment, applications, and support to reduce risk, speed up time to deployment, and ease the integration into a client's environment. Intelligent Cluster reduces the complexity of deployment with pre-integrated and interoperability-tested solutions, which are delivered, installed¹, and supported by Lenovo as end-to-end solutions.

Intelligent Cluster features industry-leading System x servers, storage, networking, software, and third-party components with which clients can choose from various technologies and design a tailored solution for the clients applications and environment. Lenovo thoroughly tests and optimizes each solution for reliability, interoperability, and maximum performance so that the system can be quickly deployed.

In Lenovo's manufacturing plant, the systems are fully assembled and integrated in the racks (including all cables). As a result, the delivery of a single rack solution requires only to connect the PDUs to the data center power, and switches uplinks to the data center network to be ready for power-on. A multiple rack solution also requires that the inter-rack cabling is done (one side of the cable being already connected, the other side being labeled with location to which it must be connected).

Lenovo applies in manufacturing a firmware stack that matches the "best-recipe" that is devised by our cluster development team for solution level interoperability.

Intelligent Cluster solutions are built, tested, delivered, installed¹, and supported by Lenovo as a single solution instead of being treated as hundreds of individual components. Lenovo provides single point-of-contact, solution-level support that includes System x and third-party components (such as those from Intel and Mellanox) to deliver maximum system availability throughout the life of the system, so clients can spend less time maintaining systems.

Although open systems can be racked, cabled, configured, and tested by users, we encourage clients to evaluate the benefits of having Lenovo integrate and test the system before delivery. We also suggest contacting the Lenovo System x Enterprise Solution Services cluster enablement team to speed up the commissioning after the equipment arrives.

6.2 Lenovo factory integration standards

Lenovo standards for factory integration are based on meeting a broad range of criteria, including the following criteria:

- ▶ Racks are one standard floor tile wide, and fit through 80-inch doorways.
- ▶ Cabling maintains proper bend radius for all cable types while not impeding maintenance access to any devices and allows rack doors to be closed (and locked, if required).
- ▶ All components are cabled within the rack.

Also, where practical, inter-rack cabling is connected at one end, coiled up, and temporarily fastened in the rack for shipping.

- ▶ All cables are labeled at each end, with the source and destination connection information printed on each label.

¹ The solution's installation can be performed by Business Partner or user, if wanted.

- ▶ All components are mounted inside the racks for shipping².

These standards govern the range of systems that can be factory-integrated by Lenovo and we consider them to follow best practices that are based on our design criteria. There are several alternative configuration options that are architecturally sound and based on different design criteria, as shown in the following examples:

- ▶ Not requiring rack doors: This option allows for bigger bundles of copper cables to traverse the fronts of the racks without impeding node access, which can allow more nodes per rack.
- ▶ Use of optical cabling: This option allows for smaller cable bundles, so more nodes per rack or more networks per node can be provisioned without impeding node access.
- ▶ Use of taller racks: This option allows for more chassis and nodes per rack.
- ▶ Allowing connections to switches outside the rack; for example, stacked on top of the rack.

Note: The responsibility for assuring adequate cooling of components, access for maintenance, adequate cable lengths before integration, and other considerations become the responsibility of the person or team that is configuring their own solution.

6.3 Factory testing

The Intelligent Cluster manufacturing test process is intended to meet the following objectives:

- ▶ Assure that the integrated hardware is configured and functional.
- ▶ Identify and repair any defects before or that were introduced by the integrated rack assembly process.
- ▶ Validate the complete and proper function of a system when configured to a customer's order specifications.
- ▶ Apply the current released Intelligent Cluster best-recipe, which includes lab-tested firmware levels for all servers, adapters, and devices.

The following tasks are typical of the testing that is performed by Lenovo at the factory. Other testing might be done based on unique hardware configurations or client requirements:

- ▶ All servers are fully tested as individual units before rack integration.
- ▶ After the components are installed in the rack, there is a post assembly inspection to assure that they are installed and positioned correctly.
- ▶ Lenovo configures all switches; terminal servers; keyboard, video, and mouse (KVM) over IP units, and other devices with IP addresses and host names to allow for communication and control.

These components can be set by using client-supplied information or to a default scheme. For more information about the default scheme, see this website:

http://download.boulder.ibm.com/ibmdl/pub/systems/support/system_x_pdf/intelligent_cluster_factory_settings_102411.pdf

² In rare cases, some components might ship outside of the racks if their location within a rack might lead to the rack tilting during shipment.

- ▶ Power redundancy testing

If there is a client-provided redundant power domain scheme, Lenovo tests with one domain that is powered down, then the opposite domain that is powered down, to ensure that all devices with redundant power feeds stay powered on.

Otherwise, remove and restore power from each PDU to assure all devices with redundant power feeds stay powered on.

- ▶ Flash all servers, adapters, and devices to current Lenovo development-provided best recipe.

- ▶ From a cluster management node, discover servers and program their integrated management module (IMM). This configuration allows for remote control of the computational infrastructure.

- ▶ Serial (console) over LAN (SoL) setup

Unless the cluster has terminal servers, SoL is configured and tested.

- ▶ Set up RAID arrays on local disks as defined by the client or client architect, or per Intelligent Cluster best practices.

- ▶ Set up shared storage devices, configure arrays with all disks present, and create and initialize logical disks to verify functionality.

- ▶ Install (for nodes with hard disk drives), or push out (for diskless nodes) an operating system to all servers to verify functionality of the following components:

- Server hardware

- Ethernet, InfiniBand, and Fibre Channel switches, terminal servers, and KVM switches

- Server configuration correctness, including memory configurations, correct CPU types, storage, and RAID

- ▶ Perform High-Performance Linpack (HPL) benchmarking on the system. This testing ensures that the following conditions are met:

- CPU and memory are stressed to the extent that is commonly found in production environments.

- Thermal stress testing is performed.

- Interconnect networks are exercised. HPL is run over the high bandwidth, low latency network. This testing is performed in groups over leaf and edge switches or at the chassis level. If no high-speed network is available, manufacturing performs a node level test.

As an added benefit, clients can see a performance baseline measurement, if requested.

Note: The HPL benchmark is run with open source libraries and no tuning parameters. Optimization is recommended for those users who are interested in determining maximum system performance.

The Intelligent Cluster testing is performed with the goal of detecting and correcting all system issues before the cluster is shipped so that the time from installation to production is minimized. This testing is meant to verify the hardware and cabling only, on a per rack basis. Any software installations require other Cluster Enablement Team, Lenovo System x Enterprise Solution Services, or third-party services.

Note: All servers with local disks have boot sectors that are wiped before they are shipped. All shared storage devices have test arrays deconstructed.

6.4 Documentation provided

This section describes the documentation Lenovo provides with the hardware. Lenovo manufacturing uses the extreme Cloud Administration Toolkit (xCAT) to set up and test systems. xCAT is also used to document the manufacturing that is set up. Because it is a popular administration utility, the xCAT table files are included with the system. The team that is setting up the cluster at the client site then uses these files in commissioning a cluster.

The following documentation is also provided in a binder with each rack and on a CD with each rack:

- ▶ MAC addresses of the Ethernet interfaces
- ▶ Machine type, model number, and serial number of devices
- ▶ Firmware levels: For servers, these levels include:
 - UEFI version
 - IMM version
 - On system diagnostics version
- ▶ Memory per server
- ▶ CPU type per server
- ▶ Proof that an operating system was running on each server, reporting the output of the `uname -r` command for each node
- ▶ PCI adapter list for each server by using the `lspci` command
- ▶ Configuration files for switches

6.4.1 HPLinpack testing results: Supplied on request

The results of the HPLinpack testing can be provided to clients, if requested. The documentation includes the following components:

- ▶ A listing of the software stack that is used, for example:
 - HPL-2.0 with the Intel timestamping patch
 - OpenBLAS
 - Mellanox OFED
 - gcc
 - MVAPICH2
 - RHEL
- ▶ A run summary, which lists GFLOPS and run time
- ▶ The detailed text output from the HPL benchmark
- ▶ Output graph with the following axis, as shown in Figure 6-1 on page 107:
 - Left y-axis = Gflops
 - Bottom x-axis = Percentage completed
 - Top x-axis = Wall clock
 - Right y-axis = Efficiency

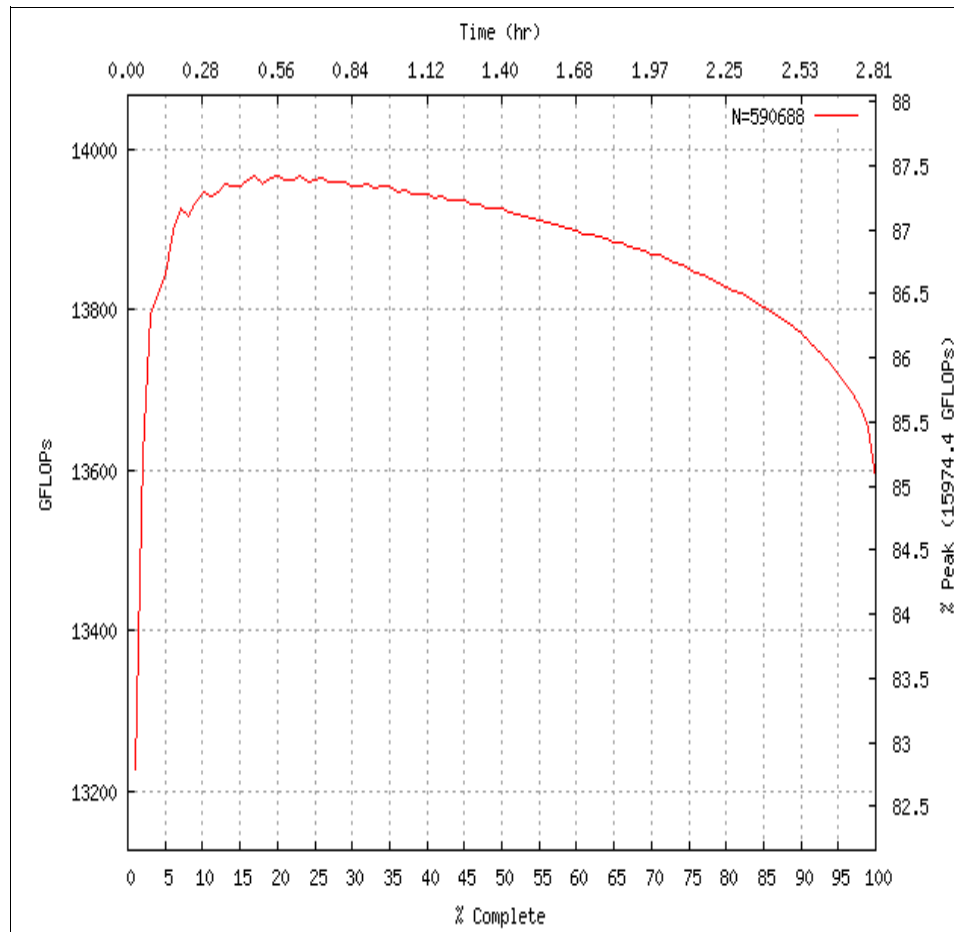


Figure 6-1 Example of HPLinpack output graph

Managing a NeXtScale environment

This chapter describes the available options for managing a NeXtScale System environment.

We describe the management capabilities and interfaces that are integrated in the system, and the middleware and software layers that are often used to manage collections of systems.

This chapter includes the following topics:

- ▶ 7.1, “Managing compute nodes” on page 110
- ▶ 7.2, “Managing the chassis” on page 123
- ▶ 7.3, “ServeRAID C100 drivers: nx360 M4” on page 148
- ▶ 7.4, “Integrated SATA controller: nx360 M5” on page 148
- ▶ 7.5, “VMware vSphere Hypervisor” on page 148
- ▶ 7.6, “eXtreme Cloud Administration Toolkit” on page 149

7.1 Managing compute nodes

The NeXtScale System compute nodes include local and remote management capabilities.

Local management capabilities are provided through the keyboard, video, mouse (KVM) connector on the front of the server. By using the console breakout cable that is included with the chassis, you can directly connect to the server console and attach USB storage devices.

Remote management capabilities are provided through the Integrated Management Module II (IMM2). IMM2 also provides advanced service control, monitoring, and alerting functions.

By default, the nx360 compute nodes include IMM2 Basic; however, if more functionality is required, the IMM2 can be upgraded to IMM2 Standard or to IMM2 Advanced with Feature on Demand (FoD) licenses.

7.1.1 Integrated Management Module II

The Integrated Management Module II (IMM2) on the NeXtScale nodes are compliant with Intelligent Platform Management Interface version 2.0 (IPMI 2.0). By using IPMI 2.0, administrators can manage a system remotely out of band, which means that a system can be managed independently of the operating system or in the absence of an operating system, even if the monitored system is not powered on.

IPMI also functions when the operating system is started and offers enhanced features when used with system management software. The nodes respond to IPMI 2.0 commands to report operational status, retrieve hardware logs, or issue requests. The nodes can alert by way of the simple network management protocol (SNMP) or platform event traps (PET).

The IMM2 can be accessed and controlled through any of the following methods:

- ▶ Command-line interface (CLI): Telnet or Secure Shell (SSH)
- ▶ Web interface (if IMM2 Standard and Advanced FoD is provisioned)
- ▶ IPMI 2.0 (local or remote)
- ▶ The Advanced Settings Utility (ASU)
- ▶ SNMP v1 and v3

Figure 7-1 shows the available IMM2 access methods.

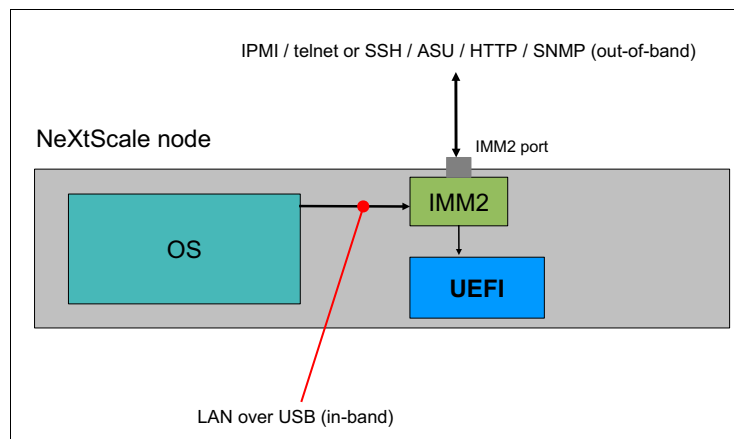


Figure 7-1 IMM2 access methods

Remote access to the IMM2 is provided by an Ethernet connection, either a port that is shared with the operating system, or a port that is dedicated to the IMM2. Consider the following points:

- The nx360 M4 has a dedicated port that allows direct access to the IMM and provides a separate physical connection, as shown in Figure 7-2.

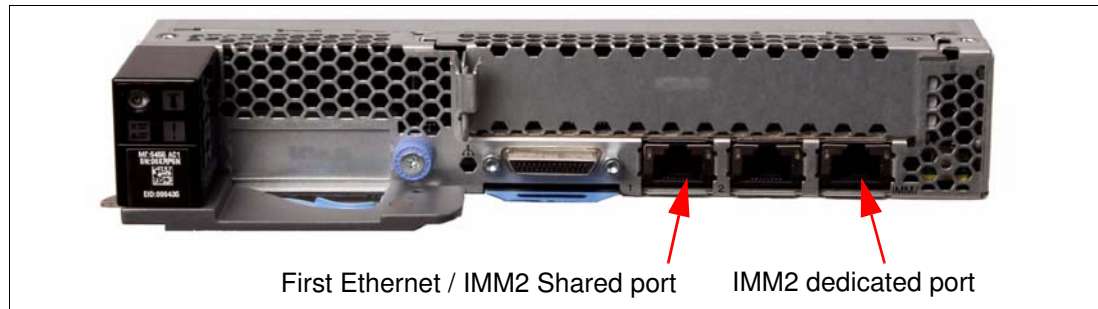


Figure 7-2 nx360 M4 IMM2 dedicated and shared ports

- The nx360 M5 has an optional dedicated port, as shown in Figure 7-3. The IMM can be configured to be accessed through the first on-board 1 Gb Ethernet port, which results in less cabling and a shared physical link to the IMM and the Ethernet port.

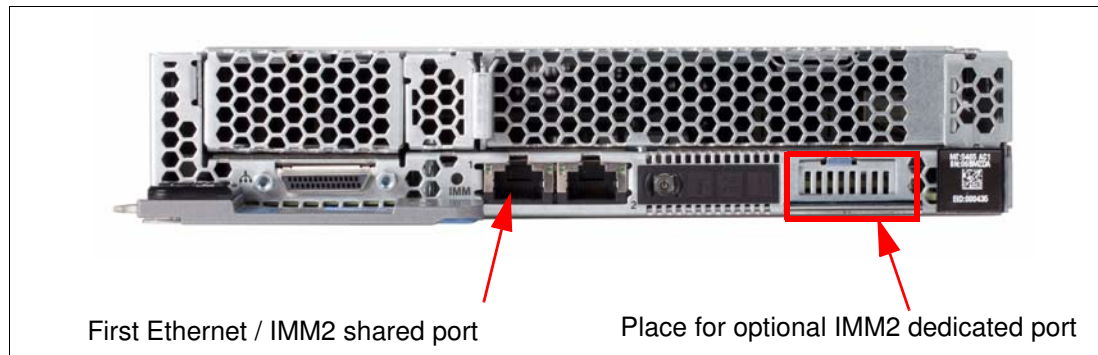


Figure 7-3 nx360 M5 IMM2 dedicated and shared ports

Dedicated and shared mode are exclusive. If shared mode is selected, the IMM2 dedicated port becomes disabled and the IMM2 can be accessed via the first Ethernet interface only. The way IMM2 is accessed can be selected manually through F1 UEFI setup menu or by using the ASU, which allows modifications to system settings through a CLI remotely or locally to the node. Figure 7-4 shows the UEFI setup to change Dedicated or Shared access interface for IMM2. For more information about ASU, see 7.1.3, “ASU” on page 203.

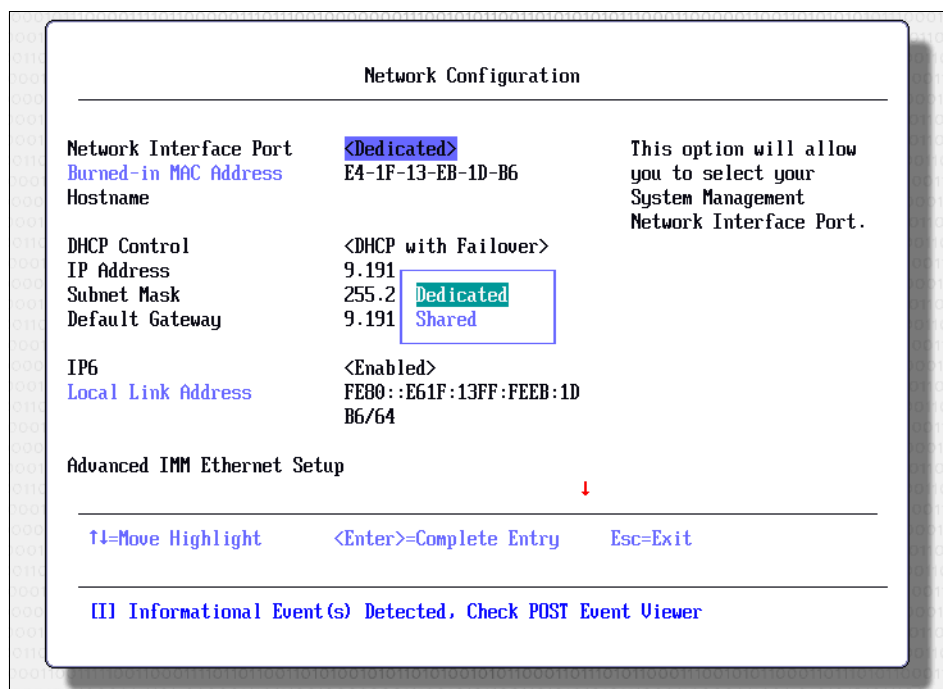


Figure 7-4 F1 UEFI menu at system start to configure IMM access

Figure 7-5 shows how the IMM2 access port is set by using the ASU tool from inside the operating system.

```
> /opt/ibm/toolscenter/asu/asu64 set IMM.SharedNicMode Shared
Advanced Settings Utility version 9.41.81K
Licensed Materials - Property of IBM
(C) Copyright IBM Corp. 2007-2013 All Rights Reserved
Successfully discovered the IMM via SLP.
Discovered IMM at IP address 169.254.95.118
Connected to IMM at IP address 169.254.95.118
IMM.SharedNicMode=Shared
Waiting for command completion status.
Command completed successfully.
```

Figure 7-5 Setting the IMM2 access port via ASU

For more information about the IMM2, see the following publications:

- IMM2 User’s Guide:
<http://ibm.com/support/entry/portal/docdisplay?ln docid=MIGR-5086346>
- A white paper about transitioning to UEFI and IMM:
<http://ibm.com/support/entry/portal/docdisplay?ln docid=MIGR-5079769>

7.1.2 Unified Extensible Firmware Interface

The Unified Extensible Firmware Interface (UEFI) replaces BIOS in System x and BladeCenter servers. It is the new interface between the operating system and platform firmware. UEFI provides a modern, well-defined environment for booting an operating system and running pre-boot applications.

For more information about UEFI, see this website:

<http://www.uefi.org/home/>

UEFI provides the following improvements over BIOS:

- ▶ ASU now has complete coverage of system settings.
- ▶ On rack mount servers, UEFI settings can be accessed out-of-band by ASU and the IMM (not available on BladeCenter blades).
- ▶ Adapter configuration can move into F1 setup; for example, iSCSI configuration is now in F1 setup and consolidated into ASU.
- ▶ Elimination of beep codes: All errors are displayed by using light path diagnostics.
- ▶ DOS is not supported and does not work under UEFI.

UEFI adds the following functionality:

- ▶ Adapter vendors can add more features in their options (for example, IPv6).
- ▶ Modular design allows faster updates as new features are introduced.
- ▶ More adapters can be installed and used simultaneously; optional ROM space is much larger.
- ▶ BIOS is supported via a legacy compatibility mode.
- ▶ Provides an improved user interface.
- ▶ Replaces Ctrl key sequences with a more intuitive human interface.
- ▶ Adapter and iSCSI configuration are moved into F1 setup.
- ▶ Event logs are created that are more easily decipherable.
- ▶ Provides easier management.
- ▶ Reduces the number of error messages and eliminates outdated errors.
- ▶ A complete setup solution is provided by allowing adapter configuration function to be moved into UEFI.
- ▶ Complete out-of-band coverage by ASU simplifies remote setup.
- ▶ More functionality, better user interface, easier management for users.

For more information about the UEFI, see the IBM white paper, *Introducing UEFI-Compliant Firmware on IBM System x and BladeCenter servers*, which is available at this website:

<http://www.ibm.com/support/entry/portal/docdisplay?lnodocid=MIGR-5083207>

UEFI system settings

Many of the advanced technology options that are available in the NeXtScale nx360 M4 compute node are controlled in the UEFI system settings. These settings affect processor and memory subsystem performance regarding power consumption.

The UEFI page is accessed by pressing F1 during the system initialization process. Figure 7-4 shows the UEFI settings main panel.

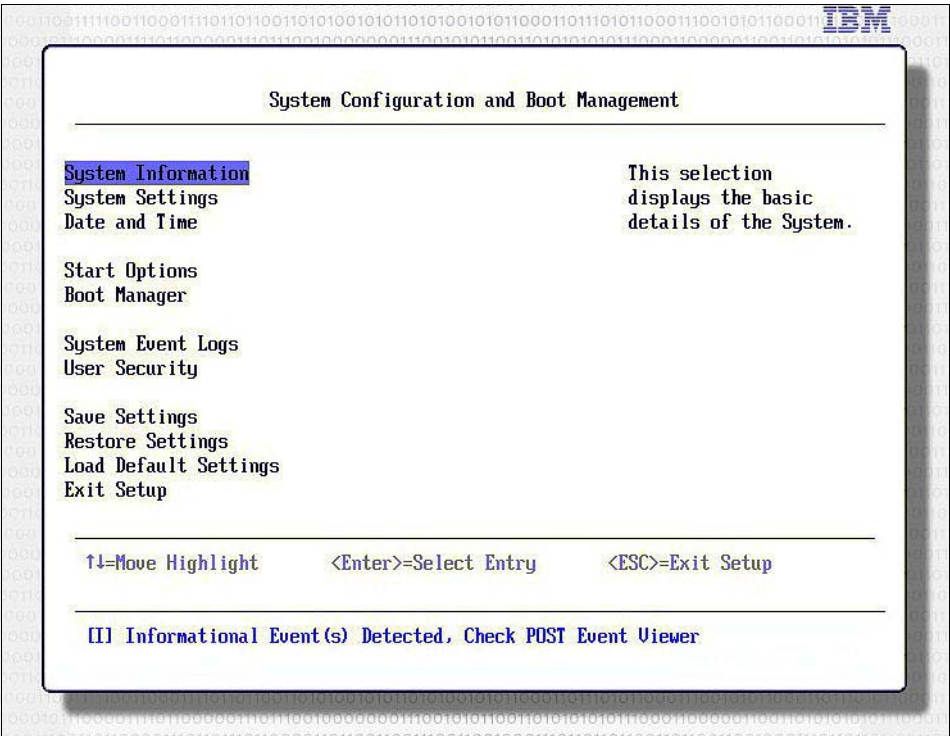


Figure 7-6 UEFI settings main panel

The compute node provides optimal performance with reasonable power usage, which depends on the operating frequency and voltage of the processors and memory subsystem.

In most operating conditions, the default settings provide the best performance possible without wasting energy during off-peak usage. However, for certain workloads, it might be appropriate to change these settings to meet specific power to performance requirements.

The UEFI provides several predefined setups for commonly wanted operation conditions. These predefined values are referred to as *operating modes*. Access the menu in UEFI by selecting **System Settings** → **Operating Modes** → **Choose Operating Mode**. You see the five operating modes from which to choose, as shown in Figure 7-5. When a mode is chosen, the affected settings change to the shown predetermined values.

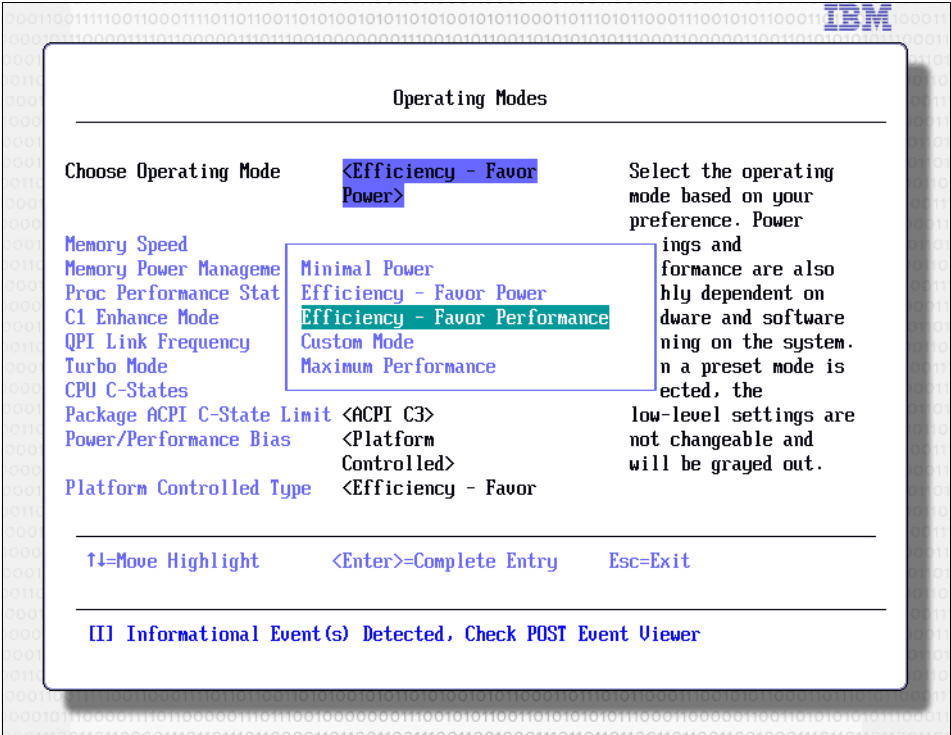


Figure 7-7 Operating modes in UEFI

We describe these modes in the following sections.

Minimal Power

Figure 7-6 shows the Minimal Power predetermined values. These values emphasize power-saving server operation by setting the processors, QPI link, and memory subsystem to a lowest working frequency. Minimal Power provides less heat and the lowest power usage at the expense of performance.

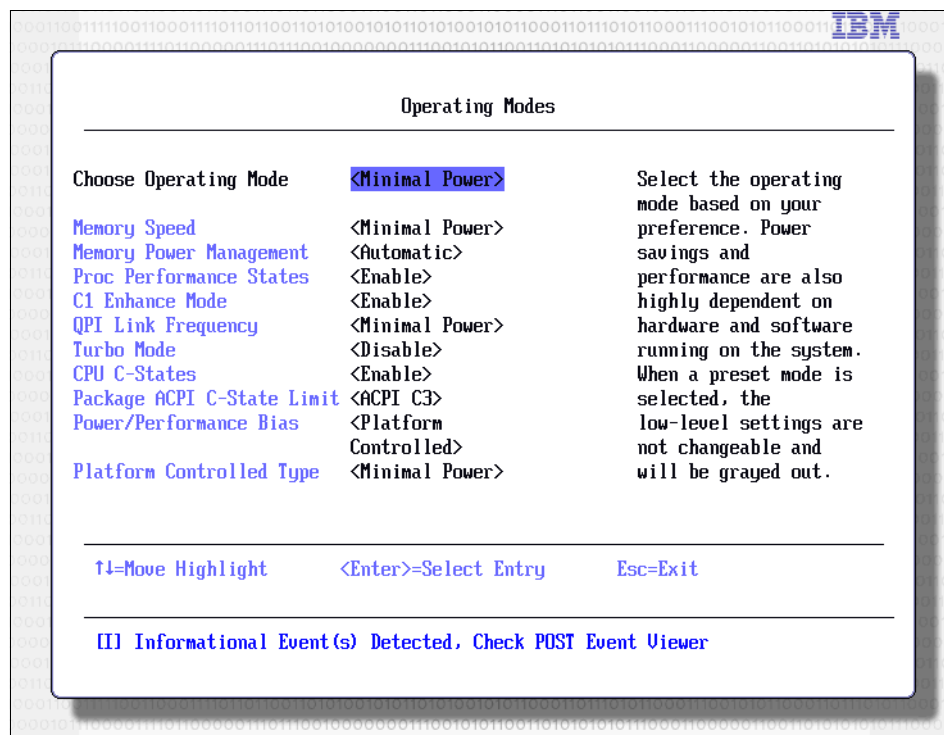


Figure 7-8 UEFI operation mode: Minimal Power

Efficiency - Favor Power

Figure 7-7 shows the Efficiency - Favor Power predetermined values. These values emphasize power-saving server operation by setting the processors, QPI link, and memory subsystem to a balanced working frequency. Efficiency - Favor Power provides more performance than Minimal Power, but favors power usage.

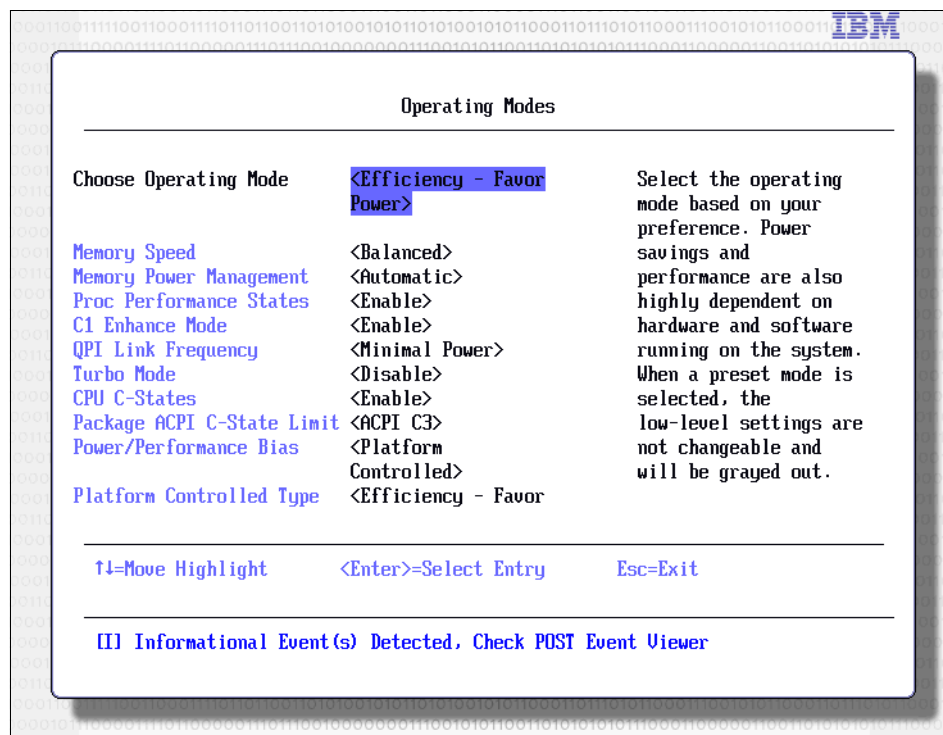


Figure 7-9 UEFI operation mode: Efficiency - Favor Power

Efficiency - Favor Performance

Figure 7-8 shows the Efficiency - Favor Performance predetermined values. These values emphasize performance server operation by setting the processors, QPI link, and memory subsystem to a high working frequency.

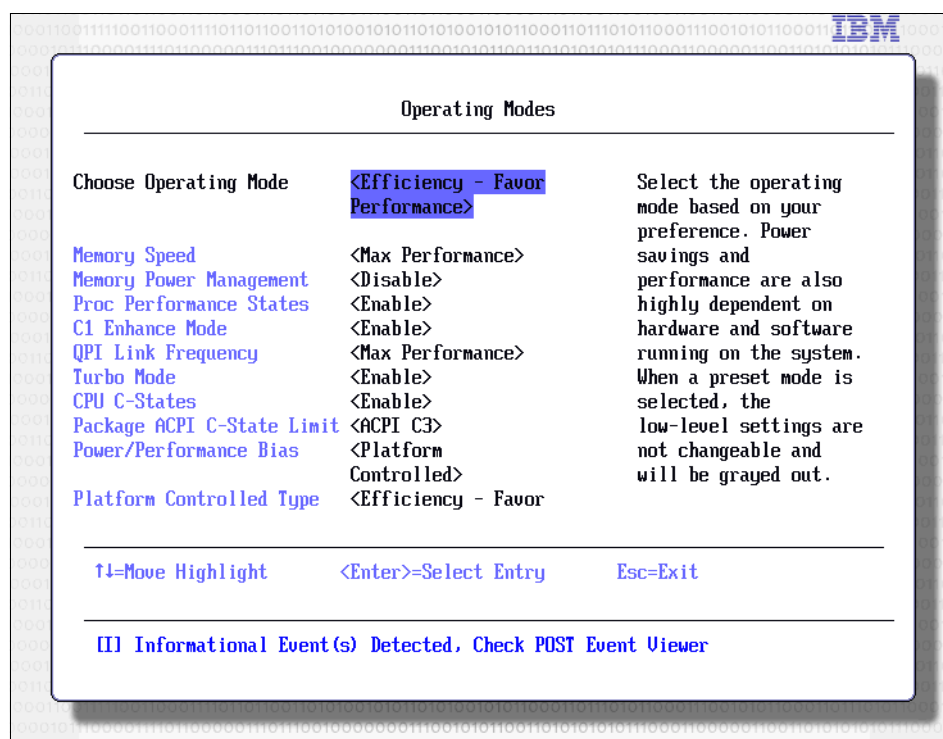


Figure 7-10 UEFI operation mode: Efficiency - Favor Performance

Custom Mode

By using Custom Mode, users can select the specific values that they want, as shown in Figure 7-9. The recommended factory default setting values provide optimal performance with reasonable power usage. However, with this mode, users can individually set the power-related and performance-related options.

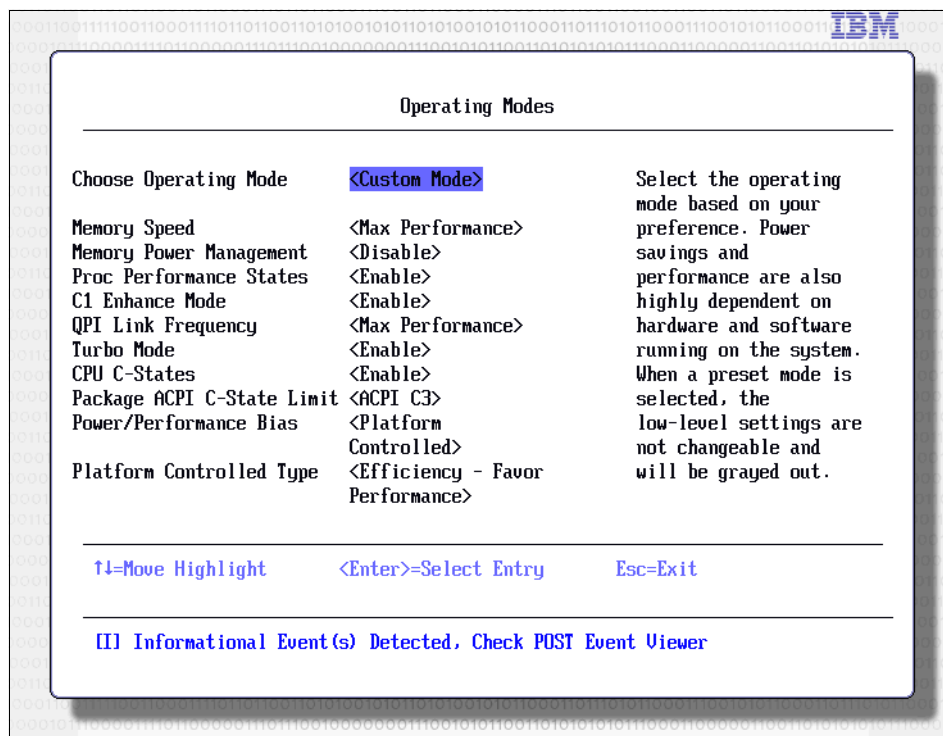


Figure 7-11 UEFI operation mode: Custom Mode

Maximum Performance

Figure 7-10 shows the Maximum Performance predetermined values. They emphasize performance server operation by setting the processors, QPI link, and memory subsystem to a maximum working frequency and the higher C-state limit. The server is set to use the maximum performance limits within UEFI. These values include turning off several power management features of the processor to provide the maximum performance from the processors and memory subsystem.

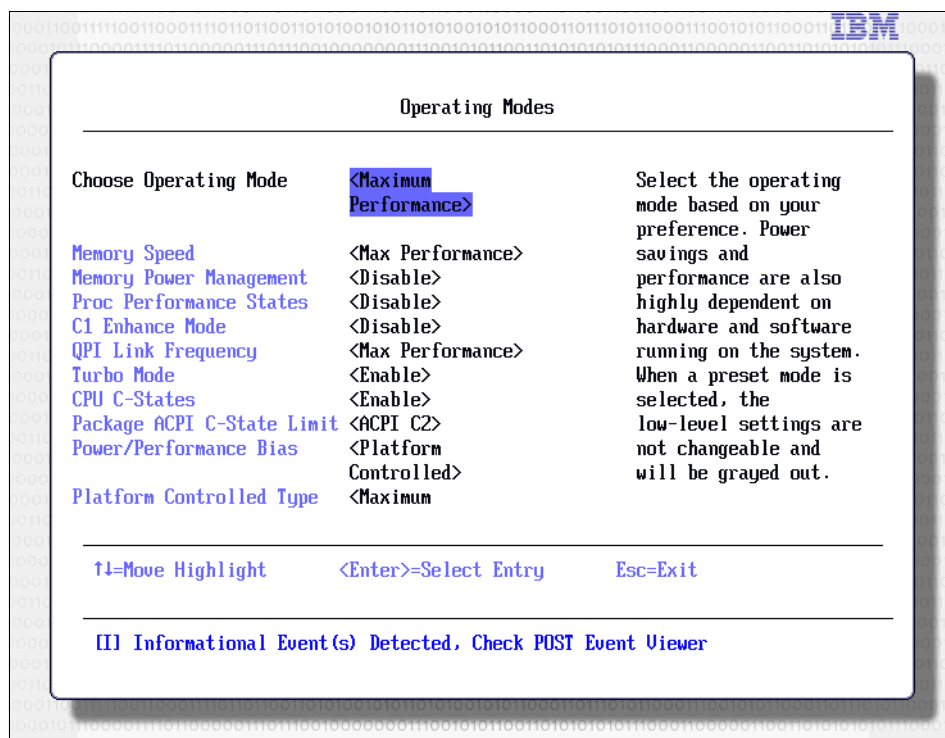


Figure 7-12 UEFI operation mode: Maximum Performance

Performance-related individual system settings

The UEFI default settings are configured to provide optimal performance with reasonable power usage. Other operating modes are also available to meet various power and performance requirements. However, individual system settings enable users to fine-tune the wanted characteristics of the compute nodes.

This section describes the UEFI settings that are related to system performance. In most cases, increasing system performance increases the power usage of the system.

Processors

Processor settings control the various performance and power features that are available on the installed Xeon processor.

Figure 7-11 shows the UEFI Processors system settings window with the default values.

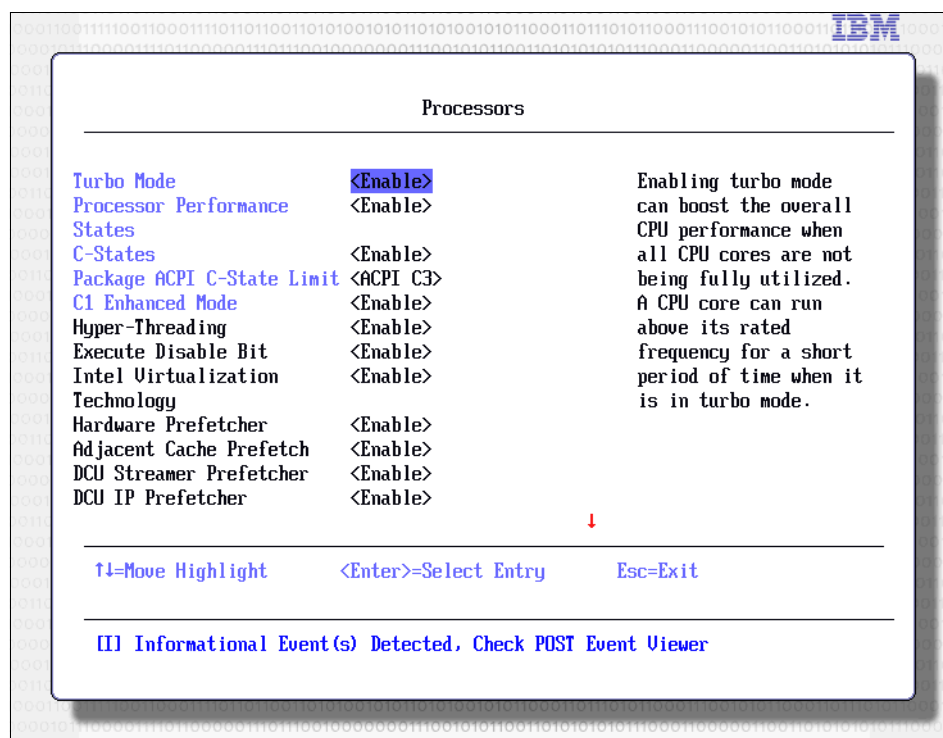


Figure 7-13 UEFI Processor system settings panel

The following processor feature options are available:

- ▶ Turbo Mode (Default: Enable)
This mode enables the processor to increase its clock speed dynamically if the CPU does not exceed the Thermal Design Power (TDP) for which it was designed.
- ▶ Processor Performance States (Default: Enable)
This option enables Intel Enhanced SpeedStep Technology that controls dynamic processor frequency and voltage changes, depending on operation.
- ▶ C-States (Default: Enable)
This option enables dynamic processor frequency and voltage changes in the idle state, which provides potentially better power savings.
- ▶ Package ACPI C-State Limit
Sets the higher C-state limit. A higher C-state limit allows the CPU to use less power when they are idle.
- ▶ C1 Enhanced Mode (Default: Enable)
This option enables processor cores to enter an enhanced halt state to lower the voltage requirement, and it provides better power savings.
- ▶ Hyper-Threading (Default: Enable)
This option enables logical multithreading in the processor so that the operating system can run two threads simultaneously for each physical core.
- ▶ Execute Disable Bit (Default: Enable)
This option enables the processor to disable the running of certain memory areas, which prevents buffer overflow attacks.

- ▶ Intel Virtualization Technology (Default: Enable)
This option enables the processor hardware acceleration feature for virtualization.
- ▶ Technology Hardware Prefetcher (Default: Enable)
This option enables the hardware prefetcher. Lightly threaded applications and some benchmarks can benefit from having it enabled.
- ▶ Adjacent Cache Prefetch (Default: Enable)
This option enables the adjacent cache line prefetch. Some applications and benchmarks can benefit from having it enabled.
- ▶ DCU Streamer Prefetcher (Default: Enable)
This option enables the stream prefetcher. Some applications and benchmarks can benefit from having it enabled.
- ▶ DCU IP Prefetcher (Default: Enable)
This option enables Instruction Pointer prefetcher. Some applications and benchmarks can benefit from having it disabled.

Memory

The Memory settings window provides the available memory operation options, as shown in Figure 7-12.

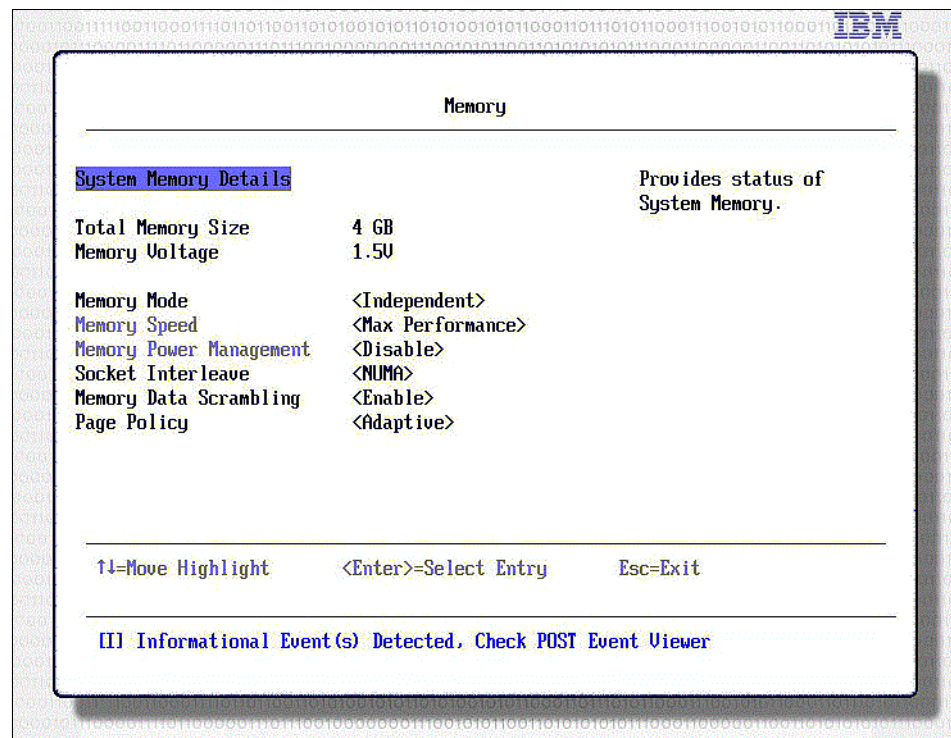


Figure 7-14 UEFI Memory system settings panel

The following memory feature options are available:

- ▶ Memory Mode (Default: Independent)
This option selects memory mode at initialization. Independent, mirroring, or sparing memory mode can be selected.

- ▶ **Memory Speed (Default: Max Performance)**
This option sets the following operating frequency of the installed DIMMs:
 - Minimal Power provides less performance for better power savings. The memory operates at the lowest supported frequency.
 - Power Efficiency provides the best performance per watt ratio. The memory operates one step under the rated frequency.
 - Max Performance provides the best system performance. The memory operates at the rated frequency.
- ▶ **Memory Power Management (Default: Disabled)**
This option sets the memory power management. Disable provides maximum performance at the expense of power.
- ▶ **Socket Interleave (Default: NUMA)**
This option sets NUMA or Non-NUMA system behavior. When NUMA is selected, memory is not interleaved across processors whereas Non-NUMA memory is interleaved across processors.
- ▶ **Memory Data Scrambling (Default: Enabled)**
This option enables a memory data scrambling feature to further minimize bit-data errors.
- ▶ **Page Policy (Default: Adaptive)**
This option determines the following Page Manager Policy in evaluating memory access:
 - Closed: Memory pages are closed immediately after each transaction.
 - Open: Memory pages are left open for a finite time after each transaction for possible recurring access.
 - Adaptive: Use Adaptive Page Policy to decide the memory page state.

7.1.3 ASU

By using the IBM ASU tool, users can modify firmware settings from the command line on multiple operating-system platforms.

You can perform the following tasks by using the utility:

- ▶ Modify selected basic input/output system (BIOS) CMOS settings without restarting the system to access F1 settings.
- ▶ Modify selected baseboard management controller setup settings.
- ▶ Modify selected Remote Supervisor Adapter and Remote Supervisor Adapter II setup settings.
- ▶ Modify selected settings in the integrated management module IMM-based servers for the IMM firmware and IBM System x Server firmware. IBM System x Server Firmware is the IBM implementation of UEFI.
- ▶ Modify a limited number of vital product data (VPD) settings on IMM-based servers.
- ▶ Modify iSCSI boot settings. To modify iSCSI settings with the ASU, you must first manually configure the values by using the server Setup utility settings on IMM-based servers.
- ▶ Connect remotely to set the listed firmware types settings on IMM-based servers. Remote connection support requires accessing the IMM external port over a LAN.

The ASU utility and documentation are available at this website:

<http://ibm.com/support/entry/portal/docdisplay?ln docid=MIGR-5085890>

Note: By using the ASU utility, you can generate a UEFI settings file from a system. A standard settings file is not provided on the IBM support site.

By using the ASU utility command line, you can read or modify firmware settings of a single node. The tool can be used from inside the node to modify node settings or to change a remote node through its IMM interface. When local to the node, ASU configures the in-band LAN over the USB interface and performs the wanted action.

When a node is accessed remotely, the IP address, user name, and password of the remote IMM must be provided. Figure 7-13 shows a command-line example that sets the IMM hostname value.

```
export PATH="/opt/ibm/toolscenter/asu/:$PATH"
asu64 set IMM.HostName1 <new_imm_hostname> --host <imm_ip_address>
      --user USERID --password PASSWORD
```

Figure 7-15 Setting the IMM2 host name that uses ASU

When the same value must be applied to several nodes, the **asu64** command must be run for each node that requires the setting. The latest releases of xCAT cluster management software provide a new tool, Parallel ASU, with which you can run the ASU tool to several nodes in parallel. xCAT cluster management capabilities and Parallel ASU help the IT administrator to perform changes to its cluster.

For more information about xCAT, see 7.6, “eXtreme Cloud Administration Toolkit” on page 149.

7.1.4 Firmware upgrade

Upgrading the firmware of NeXtScale System servers uses the same tools as other System x servers. The following tools are available to update the firmware of the NeXtScale nx360 M4 compute node, such as UEFI, IMM, or drivers:

- Stand-alone updates

Firmware updates can be performed online and offline. Offline updates are released in a bootable diskette (.img) or CD-ROM (.iso) format. Online updates are released for Windows (.exe), Linux, and VMware (.sh). The online updates are run under from the command line. They are scriptable and provided with XML metafiles for use with UpdateXpress System Packs.

Updates are available as Windows (.exe) or Linux (.bin) files, and are applied locally through the operating system, or remotely through the IMM2 Ethernet interface.

- UpdateXpress System Packs

UpdateXpress System Packs are a tested set of online updates that are released together as a downloadable package when products are first made available, and quarterly thereafter. The UpdateXpress System Packs are unique to each model of server.

The latest version of UXSP and a user’s guide are available at this website:

<http://ibm.com/support/entry/portal/docdisplay?ln docid=LNVO-XPRESS>

- ▶ **Bootable Media Creator**

Lenovo offers Bootable Media Creator with which you can bundle multiple System x updates from UpdateXpress System Packs and create bootable media, such as a CD or DVD .iso image, a USB flash drive, or a file set for PXE boot. You can download the tool from this website:

<http://ibm.com/support/entry/portal/docdisplay?lnocid=LNVO-BOMC>

7.2 Managing the chassis

The NeXtScale n1200 Enclosure includes the Fan and Power Controller (FPC) that manages power supply units and fans that are at the rear of the enclosure. In contrast to the IBM BladeCenter or IBM Flex System enclosures, the NeXtScale n1200 Enclosure does not contain a module that allows managements operation to be performed on the installed nodes. The FPC module is kept simple, and as with the iDataPlex system, compute nodes are managed through their IMM2 interface rather than a management module in the chassis.

The FPC module provides information regarding power usage at chassis or node level, fan speed, and allows the setting of specific power and cooling policies. The FPC module also adds a few management capabilities of the compute nodes that are installed in the chassis.

The FPC module includes the following features:

- ▶ Power usage information at the chassis, node, power supply units (PSU), and fan levels.
- ▶ PSU and fan status information.
- ▶ Configures wanted redundancy modes for operation (non-redundant, N+1, N+N, and oversubscription).
- ▶ Reports fan speed, fan status, and allows acoustic modes to be set.
- ▶ Defines a power cap policy at chassis or node level.

For more information about features and functional behavior, see 3.8, “Fan and Power Controller” on page 30.

Managing and configuring the FPC module can be done through a basic web browser interface that is accessed remotely by its Ethernet connection or remotely through the IPMI interface it provides.

The following sections describe how FPC module can be managed and configured from both interfaces. Although a web browser interface is intended to manage a single chassis, the IPMI interface can be used to develop wrappers that enclose IPMI commands to manage multiple chassis in the network:

- ▶ 7.2.1, “FPC web browser interface” on page 123
- ▶ 7.2.2, “FPC IPMI interface” on page 140

7.2.1 FPC web browser interface

The FPC module web interface can be accessed by the Ethernet connection at the rear of the chassis. The web interface provides a graphical and easy to manage method to configure a single chassis. However, when multiple chassis are managed, the best option is to remotely manage them through the IPMI over LAN interface (for more information, see 7.2.2, “FPC IPMI interface” on page 140).

Complete the following steps to access the FPC web interface:

1. Browse to the FPC interface URL that is defined for your FPC module. By default, the module is configured with the static IP address 192.168.0.100/24.
2. After the login window opens, enter the user name and password, as shown in Figure 7-14. The default user ID is USERID and default password is PASSWORD (where the 0 in the password is the zero character and not an uppercase letter o.)

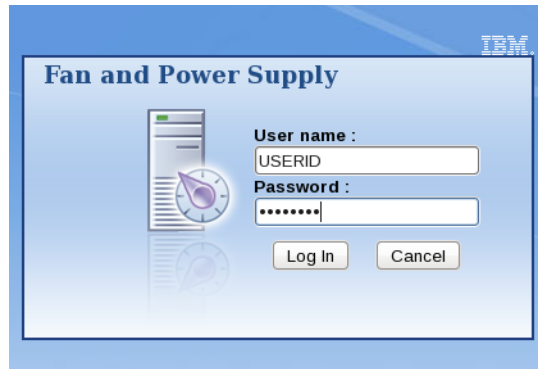


Figure 7-16 Fan and Power Controller log in page

3. Click **Log in**.

After you are logged in, the main page shows the following main functions on the left side of the page, as shown in Figure 7-15 on page 125:

- ▶ **Summary:** Displays the enclosure overall status and information. It introduces the chassis front view and rear view components and provides the status of the components (compute nodes, power supply units, fans, and so on).
- ▶ **Power:** Provides the power information about the different enclosure elements and allows the configuration of power supply redundancy modes, power capping or saving policies, and power restore policies.
- ▶ **Cooling:** Provides information about fan speed and allows the acoustic mode to be configured.
- ▶ **System Information:** Shows fixed Vital Product Data information for the enclosure, midplane, and FPC module.
- ▶ **Event Log:** Displays the System Event Log (SEL) and provides an interface to back up or restore the current configuration to or from the internal USB drive.
- ▶ **Configuration:** Allows the configuration of multiple options, such as, network, SNMP traps, alerts, and SMTP server.

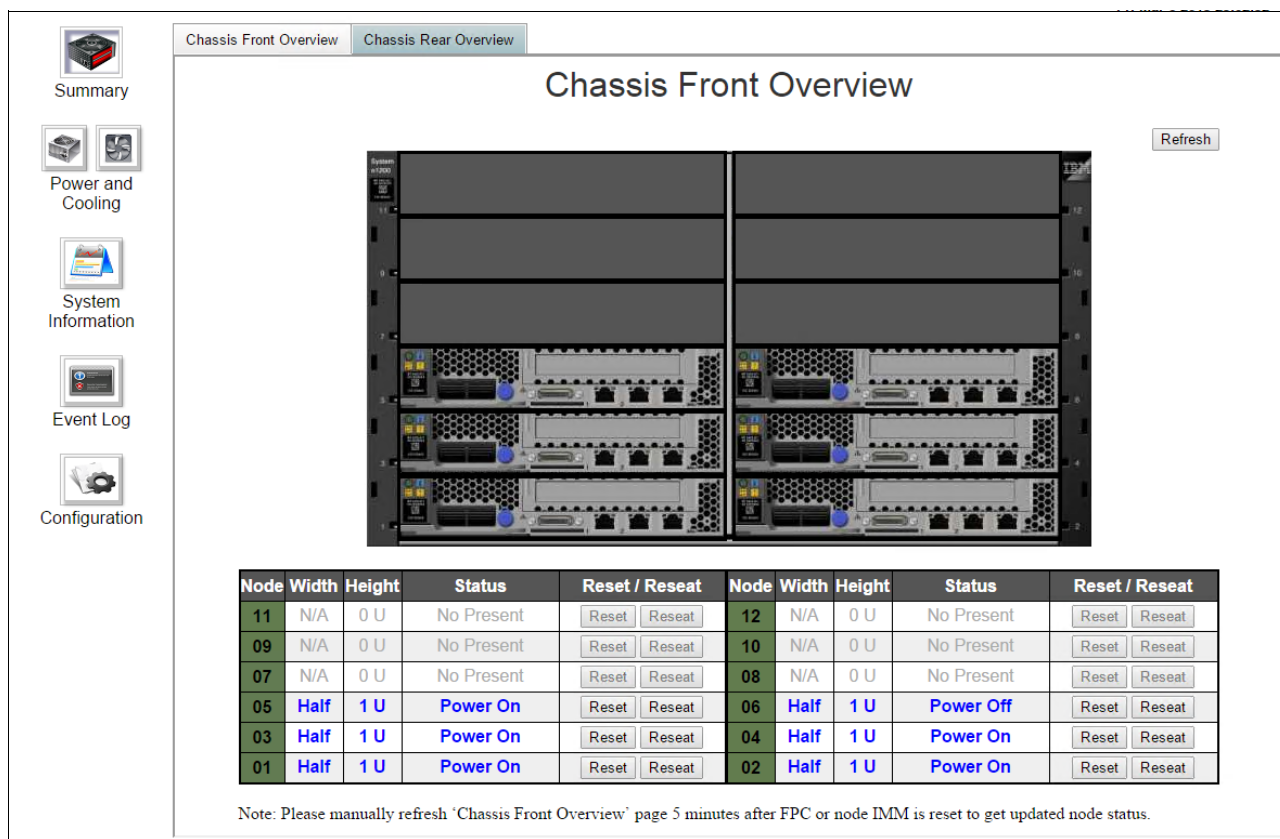


Figure 7-17 Summary front overview

The six main functions are described in the following sections:

- ▶ “Summary”
- ▶ “Power” on page 128
- ▶ “Cooling” on page 132
- ▶ “System Information tab” on page 133
- ▶ “Event Log” on page 134
- ▶ “Configuration” on page 135

Summary

The Summary function displays the enclosure’s overall status and information. There are two tabs that correspond to the front and rear of the chassis: Front Overview and Rear Overview.

Front Overview tab

As shown in Figure 7-15, the Front Overview table provides a graphical front view of the enclosure and a table that lists the status and information regarding the systems that are available in the enclosure.

Table 7-1 on page 126 lists the possible values that can appear in each column at the systems table.

Table 7-1 Front overview systems table

Column	Description
Node	Indicates slot number
Width	Possible values: <ul style="list-style-type: none"> ▶ Half: Represents a half-wide node ▶ Full: Represents a full-wide node (for future use)
Height	Node height can be 1U to 6U (for future use)
Status	Node power-on status. Possible values: <ul style="list-style-type: none"> ▶ No Present: No node is installed ▶ No Permission: Node is not granted power permission and cannot be powered on ▶ Fault: Node has a power fault and cannot be powered on ▶ Power On: Node is powered on ▶ Power Off: Node is powered off
Reset/Reseat	Used to perform virtual reset or virtual reseat: <ul style="list-style-type: none"> ▶ Virtual Reset: User can remotely reset (reboot) the IMM through the FPC. ▶ Virtual Reseat: User can remotely power cycle entire node. Reseat provides a way to emulate physical disconnection of a node. After virtual reset or reseat, node IMM takes up to two minutes to be ready.

Rear Overview tab

The Rear Overview window provides a graphical rear view of the enclosure and a table that lists the status and information regarding power supply units, system fans, and FPC module information. It displays characteristics of the available elements and a summary of health conditions, with which the system administrator can easily identify the source of a problem.

Figure 7-18 shows the rear overview window. Table 7-2 and Table 7-3 on page 210 show the possible values that can appear in each column of the tables.

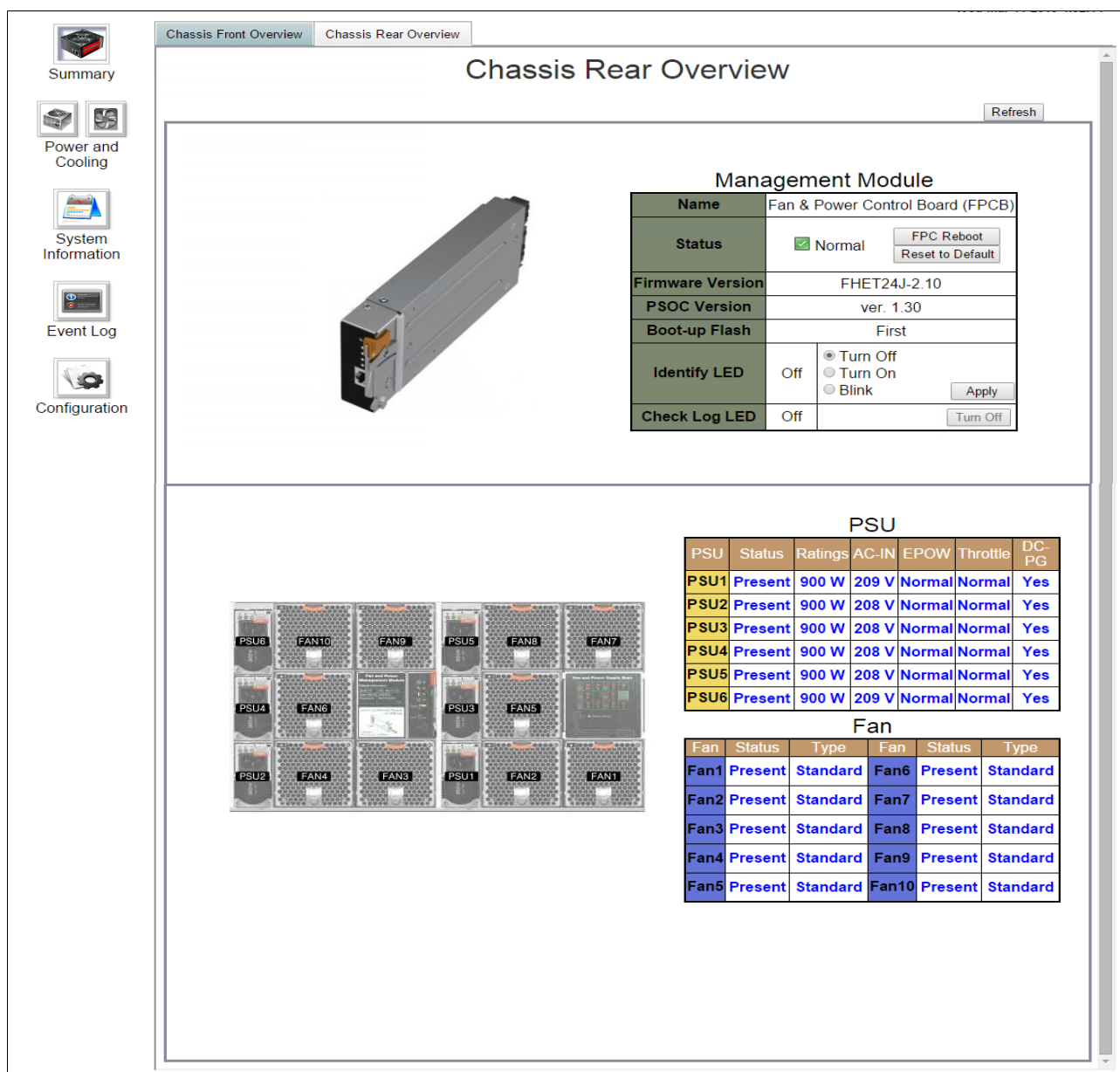


Figure 7-18 Summary rear overview

Table 7-2 Power supply unit table

Column	Description
Status	Possible values: <ul style="list-style-type: none"> ► Present: Power supply installed ► No Present: No power supply installed ► Fault: Power supply in faulty condition
Ratings	Display the DC power output rating of the power supply
AC-IN	Display the power AC input voltage rating

Column	Description
EPOW ^a	Possible values: <ul style="list-style-type: none"> ▶ Assert: Power supply is in AC lost condition ▶ Normal: Power supply is in healthy, operating condition
Throttle ^a	Possible values: <ul style="list-style-type: none"> ▶ Assert: Power supply is in over-current condition ▶ Normal: Power supply is in healthy, operating condition
DC-PG	DC power good. Possible values: <ul style="list-style-type: none"> ▶ No: Power supply is not providing the required DC power ▶ Yes: Power supply is in healthy, operating condition

a. For more information about Early Power Off Warning (EPOW) and Throttle, see 3.9, “Power management” on page 34.

Table 7-3 System fan status table

Column	Description
Status	Possible values: <ul style="list-style-type: none"> ▶ Present: fan installed ▶ No Present: no fan installed
Type	Possible values: <ul style="list-style-type: none"> ▶ Standard ▶ High performance (for future use)

Power

The Power function provides the power information about the different enclosure elements. You can configure power redundancy modes, power capping and power-saving policies, and the power restore policy to be used.

The following tabs are available and are described next:

- ▶ “Power Overview tab” on page 128
- ▶ “PSU Configuration tab” on page 129
- ▶ “Power Cap tab” on page 130
- ▶ “Voltage Overview tab” on page 132
- ▶ “Power Restore Policy tab” on page 132

Power Overview tab

As shown in Figure 7-19, the Power Overview tab provides information about the total chassis power consumption (minimum, average, and maximum of AC-in and DC-out) and a specific power information breakdown of the different systems and total system fans

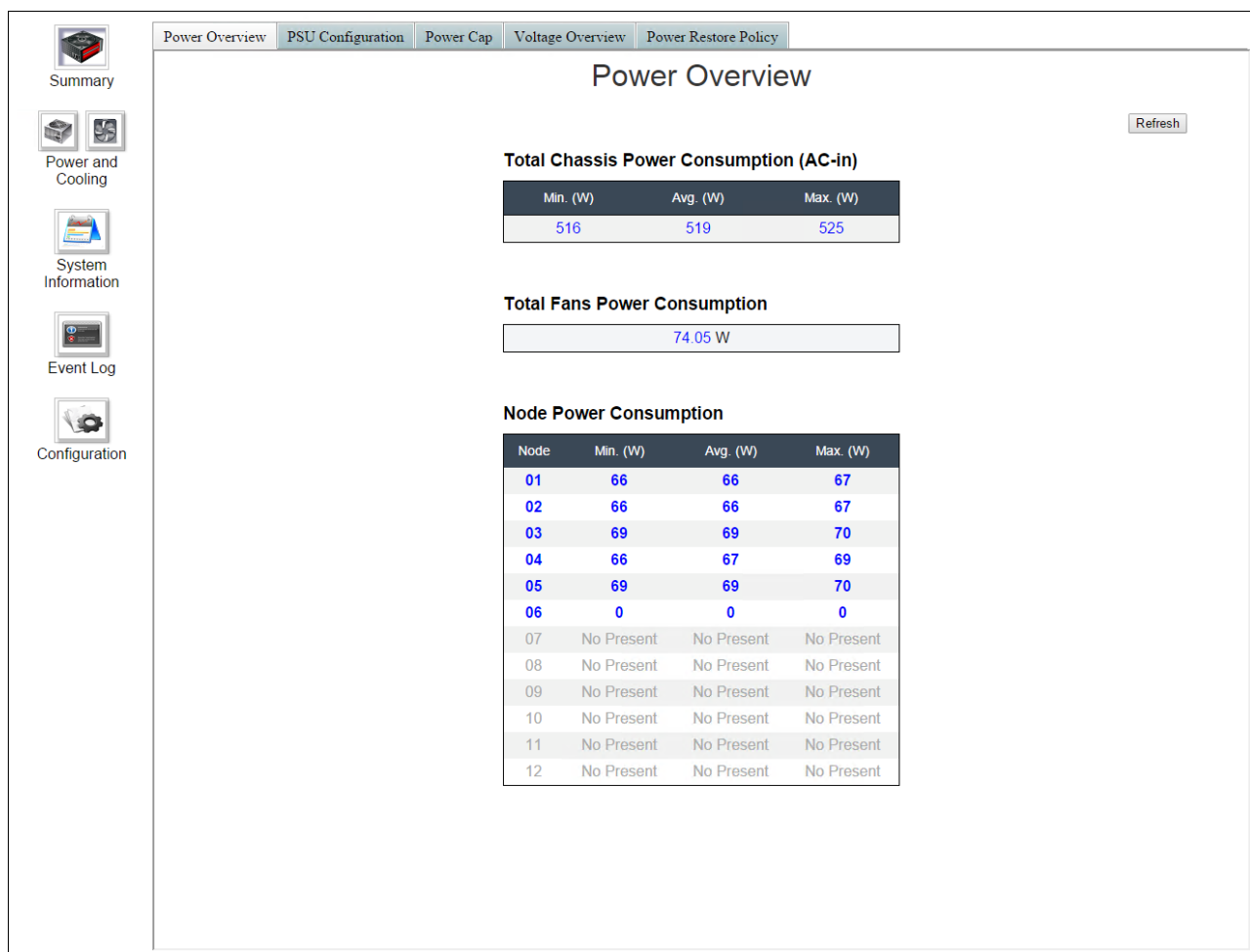


Figure 7-19 Power overview¹

The total fan power consumption can also be determined by using IPMI. The following format of the command is used:

```
ipmitool -I lanplus -H <FCP hostname or IP address> -U <userid> -P <password> raw 0x32 0x90 0x03
```

For example:

```
ipmitools -I lanplus -H 192.168.0.100 -U USERID -P PASSWORD raw 0x32 0x90 0x03
```

The output is shown in the following example:

```
00 ed 1c 00
```

Where:

- 00 in the left position is the return code (successful completion)

¹ The power consumption of the chassis depends on nodes configuration and actual load. Figure 7-19 was taken with minimum node configuration and no load. Power consumption numbers differ in each case.

- ed is the least significant byte (LSB1)
- 1c is the second least significant byte (LSB2)
- 00 in the right position is the most significant byte (MSB)

The following calculation is used for the fan power (in decimal):

$$\text{Fan power} = ((\text{MSB})_{10} \times 256^2 + (\text{LSB2})_{10} \times 256 + (\text{LSB1})_{10}) \times 0.01\text{W}$$

In our example:

- ed in hexadecimal = 237 in decimal
- 1c in hexadecimal = 28 in decimal
- Fan power = $(0 \times 256^2 + 28 \times 256 + 237) \times 0.01\text{W}$
- Fan power = $(0 + 7,168 + 237) \times 0.01\text{W}$
- Fan power = $(7,405) \times 0.01\text{W}$
- Fan power = 74.05W

PSU Configuration tab

As shown in Figure 7-18 on page 130, the PSU Configuration tab allows setting the redundancy mode for the enclosure and enable oversubscription mode, if needed.

The following redundancy modes can be selected:

- No redundancy: Compute nodes can be throttled or shutdown if any power supply is in faulty condition.
- N+1: One of the power supplies is redundant, so a single faulty power supply is allowed.
- N+N: Half of the PSUs that are installed are redundant, so the enclosure can support up to N faulty power supply units.

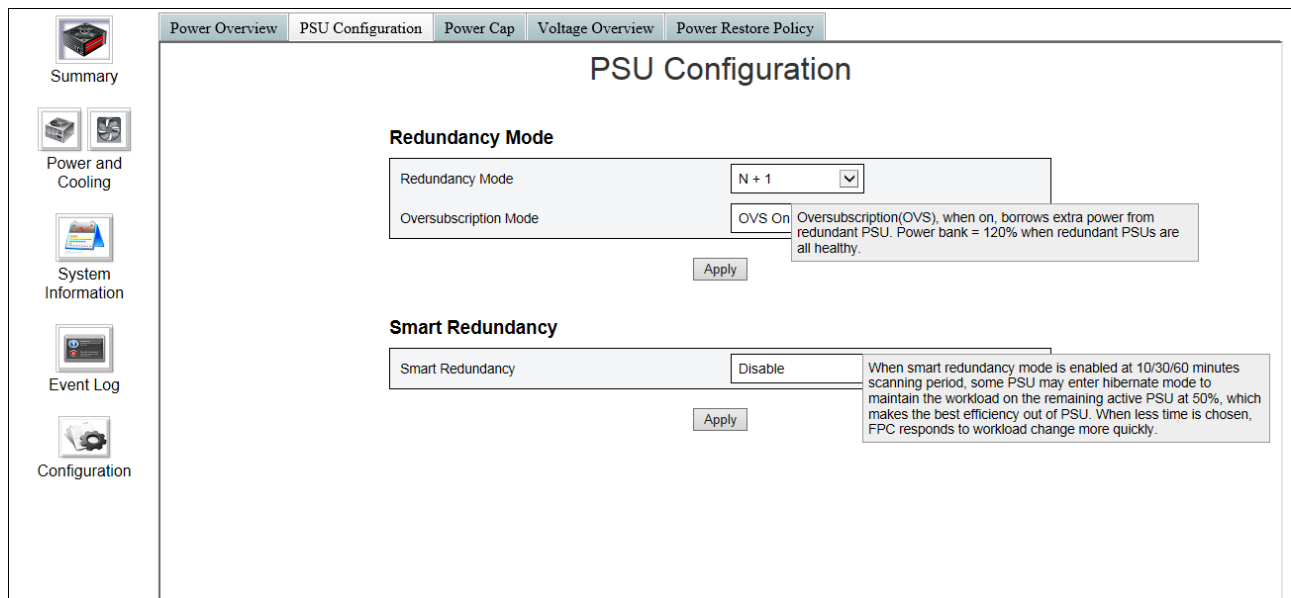


Figure 7-20 Power supply redundancy mode configuration

Oversubscription option (OVS) is only selectable when N+1 or N+N redundancy modes are enabled. When oversubscription is enabled, the enclosure allows node loads up to 120% of the redundant power budget. If a power supply fails, the surviving power supplies can sustain the extra load until the nodes are throttled to a lower power state. This mode prevents system outages while enabling higher power consumptive nodes to operate under normal and power supply fault conditions.

When smart redundancy is enabled, the FPC signals power supplies that are not required to maintain the selected redundancy mode to stop providing DC power to the chassis. This change increases the load on the remaining supplies, which improves their power efficiency.

Smart redundancy is only available with 1300 W power supplies.

The power budget that is available to grant power permission to systems that are installed in the chassis depends directly on the capacity of the power supply units that are installed, the demanding power of the nodes, the redundancy mode that was selected, and if oversubscription is enabled.

For more information, see 3.9, “Power management” on page 34.

Power Cap tab

The Power Cap tab allows setting power capping and power-saving modes at chassis or node level. Power capping and power-saving modes can be applied simultaneously.

Power capping at chassis level is selected at the drop-down menu. A range is suggested that is based on the minimum and maximum power consumption of the systems that are installed in the chassis. Any value that is not set within the suggested range is allowed; however, if it is below the minimum, it might not be reached.

Figure 7-19 shows power capping windows at chassis level. The range that is suggested is based on the aggregation of the minimum and maximum power consumption for the nodes that are installed at the chassis.

Node	Capping	Saving
Chassis	<input checked="" type="checkbox"/> 110 W (Range: 1201 W ~ 2297 W)	<input type="radio"/> Disable <input checked="" type="radio"/> Mode 1 <input type="radio"/> Mode 2 <input type="radio"/> Mode 3

Figure 7-21 Power capping at chassis level

Power capping at node level is selected via the drop-down menu. The specific node is selected in the drop-down menu that appears inside the table. Here, the suggested range is based on the minimum and maximum consumption of the node. Again, any value that is outside of the range is allowed, but it might not be reached. Figure 7-20 shows power capping windows at node level. The range that is suggested is based on the minimum and maximum power consumption for the nodes that are selected.

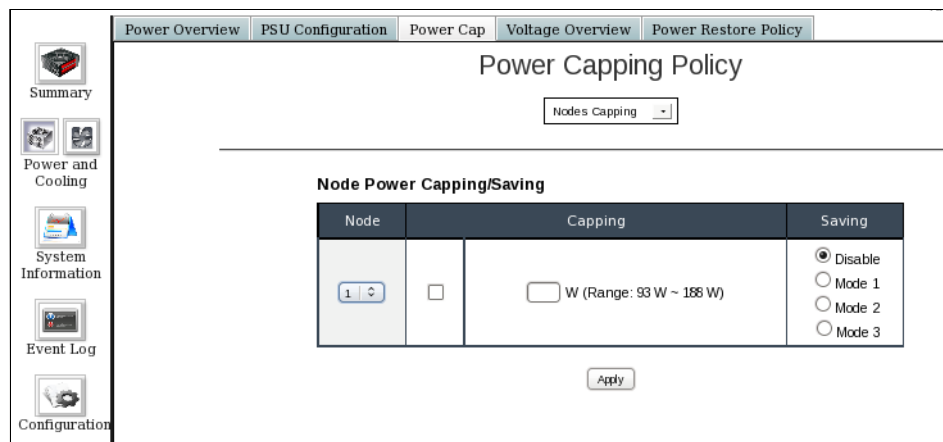


Figure 7-22 Power capping at node level

Similar to power capping, power saving can be set at chassis or node level. The following four modes can be selected:

- ▶ Disabled (default static maximum performance mode): The system runs at full speed (no throttling), regardless of the workload.
- ▶ Mode 1 (static minimum power): The system runs in a throttling state regardless of the workload. The throttling state is the lowest frequency P-state.
- ▶ Mode 2 (dynamic favor performance): The system adjusts throttling levels that are based on workload, which attempts to favor performance over power savings.
- ▶ Mode 3 (dynamic favor power): The system adjusts the throttling levels that are based on workload that is attempting to favor power savings over performance.

Voltage Overview tab

As shown in Figure 7-21, the Voltage Overview tab displays the actual FPC 12 V, 3.3 V, 5 V, and battery voltage information. If a critical threshold is reached, error log is asserted.

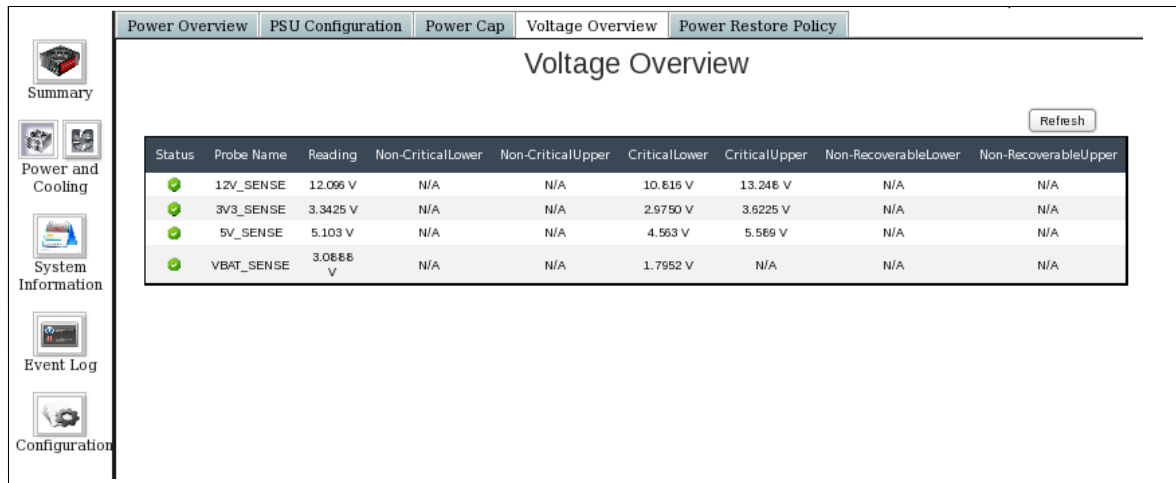


Figure 7-23 Voltage overview

Power Restore Policy tab

The Power Restore Policy tab displays and the user can set the restore policy for specific nodes. The FPC module remembers nodes that are already powered on and have power restore policy enabled. When AC is abruptly lost, it automatically turns on the nodes when AC is recovered.

To enable the restore policy on certain nodes, select the nodes and click **Apply**, as shown in Figure 7-22. The Status changes from Disable to Enable (or vice versa).

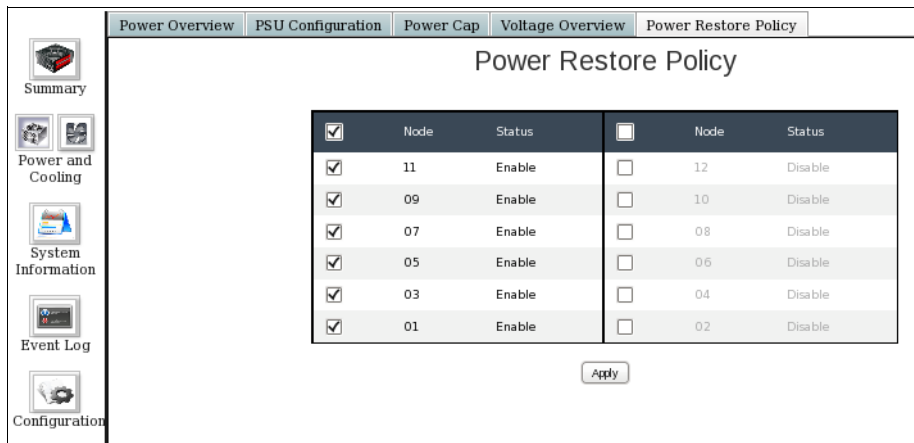


Figure 7-24 Power Restore Policy tab window

Cooling

The Cooling function provides information about fan speed for system fans and power supply unit fans. The following tabs are available:

- Cooling Overview
- PSU Fan Speed
- Acoustic Mode

These tabs are described next.

Cooling Overview tab

As shown in Figure 7-25, the Cooling Overview tab displays system fan speeds and their healthy condition. Each fan is equipped with dual motor, so A displays the primary fan motor speed and B displays the redundant fan motor speed. System fan speed normally operates in the range of 2,000 - 13,000 rpm.

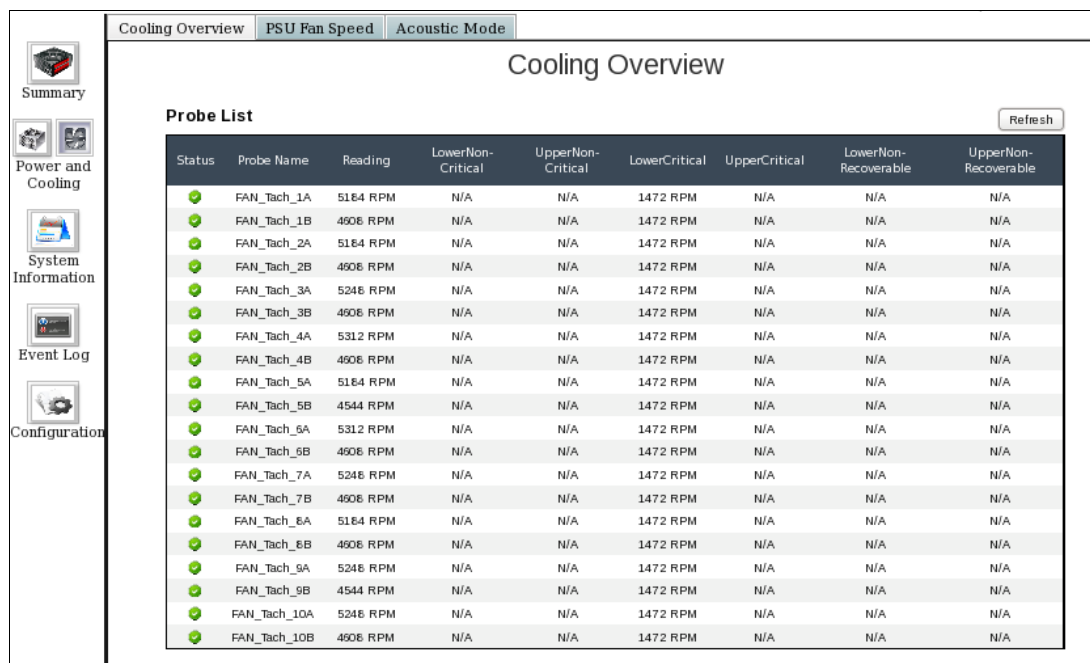


Figure 7-25 Cooling overview of system fan speeds

PSU Fan Speed tab

As shown in Figure 7-23, the PSU Fan speed tab shows the power supply fan speeds and their healthy condition. PSU fans normally operate at 5,000 - 23,000 rpm and are considered faulty when the speed falls below 3,000 rpm.

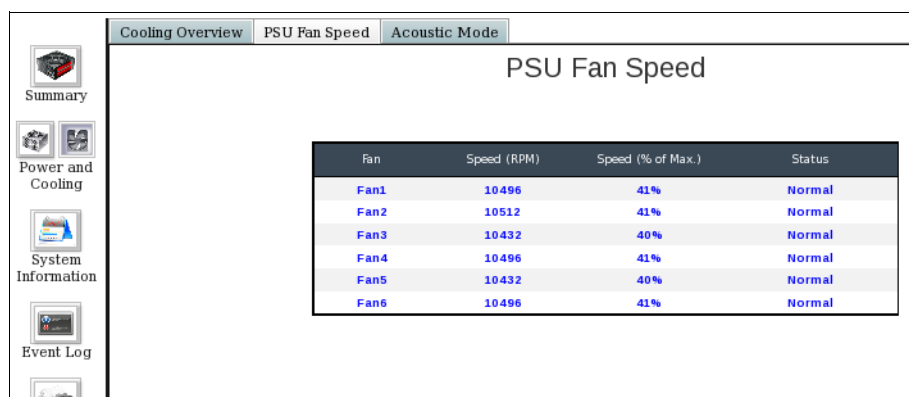


Figure 7-26 PSU Fan speed information

Acoustic Mode tab

As shown in Figure 7-27 on page 217, the Acoustic Mode is set in the Acoustic Mode tab and is intended to reduce the noise of the NeXtScale n1200 Enclosure. The following acoustic modes (which are applied to the chassis as a whole) can be selected:

- None: No acoustic mode enabled.

- ▶ Mode 1: Maximum system fan speed is 28%:
 - With 900 W power supplies, a maximum of 6.93 bels
 - With 1300 W power supplies, a maximum of 7.19 bels
- ▶ Mode 2: Maximum system fan speed is 34%:
 - With 900 W power supplies, a maximum of 7.27 bels
 - With 1300 W power supplies, a maximum of 7.42 bels
- ▶ Mode 3: Maximum system fan speed is 40%:
 - With 900 W power supplies, a maximum of 7.58 bels
 - With 1300 W power supplies, a maximum of 7.89 bels

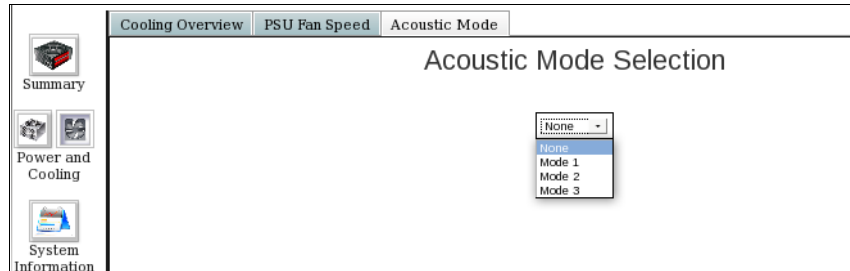


Figure 7-27 Acoustic mode selection tab

Ideally, the lowest acoustic mode setting that does not affect performance is the best choice, unless lower performance is an acceptable outcome to create a lower noise level. Acoustic mode might force nodes to be power capped or even shut down to avoid over-heating conditions.

For more information, see 3.8.5, “Acoustic mode” on page 58.

System Information tab

The System Information tab provides information about the Vital Product Data (VPD) for the chassis, the midplane, and the FPC module.

Figure 7-24, Figure 7-25, and Figure 7-26 on page 134 show the system information VPD windows for the chassis, the midplane, and the FPC.

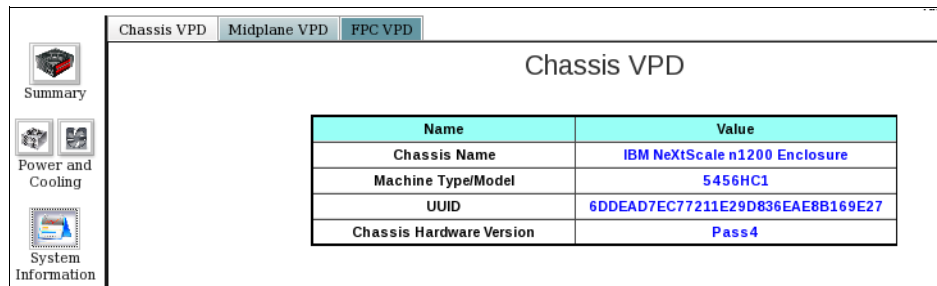


Figure 7-28 Chassis Vital Product Data window

Chassis VPD Midplane VPD FPC VPD	
Midplane VPD	
Name	Value
Midplane Name	Air Mid-plane
Card Serial Number	Y030UN 34B025
Card UUID	6DDF7C68C77211E290A26AE8B169E27
Card Hardware Version	Pass4
Card FRU Serial Number	46W2907

Figure 7-29 Midplane Vital Product Data window

Chassis VPD Midplane VPD FPC VPD	
FPC VPD	
Name	Value
FPC Name	FPC Card
Card Serial Number	Y014UN39W02B
Card UUID	F4D1FCF928B611E39577BDD9DAD2BFC1
Card Hardware Version	Pass5
Card FRU Serial Number	00Y8605

Figure 7-30 FPC Vital Product Data window

Event Log

The Event Log function displays the SEL of the chassis. Users can back up or restore user configurations to and from an internal USB.

SEL log includes information, warning, and critical events that are related to the chassis. The maximum of log entries is 512. When the log is full, you must clear it to allow new entries to be saved. When the log reaches 75%, a warning event is reported via SNMP.

You can clear the log by clicking **Clear Log** (as shown in Figure 7-27) or by running the following IPMI command:

```
ipmitool -I lanplus -U USERID -P PASSWORD -H 192.168.0.100 sel clear
```

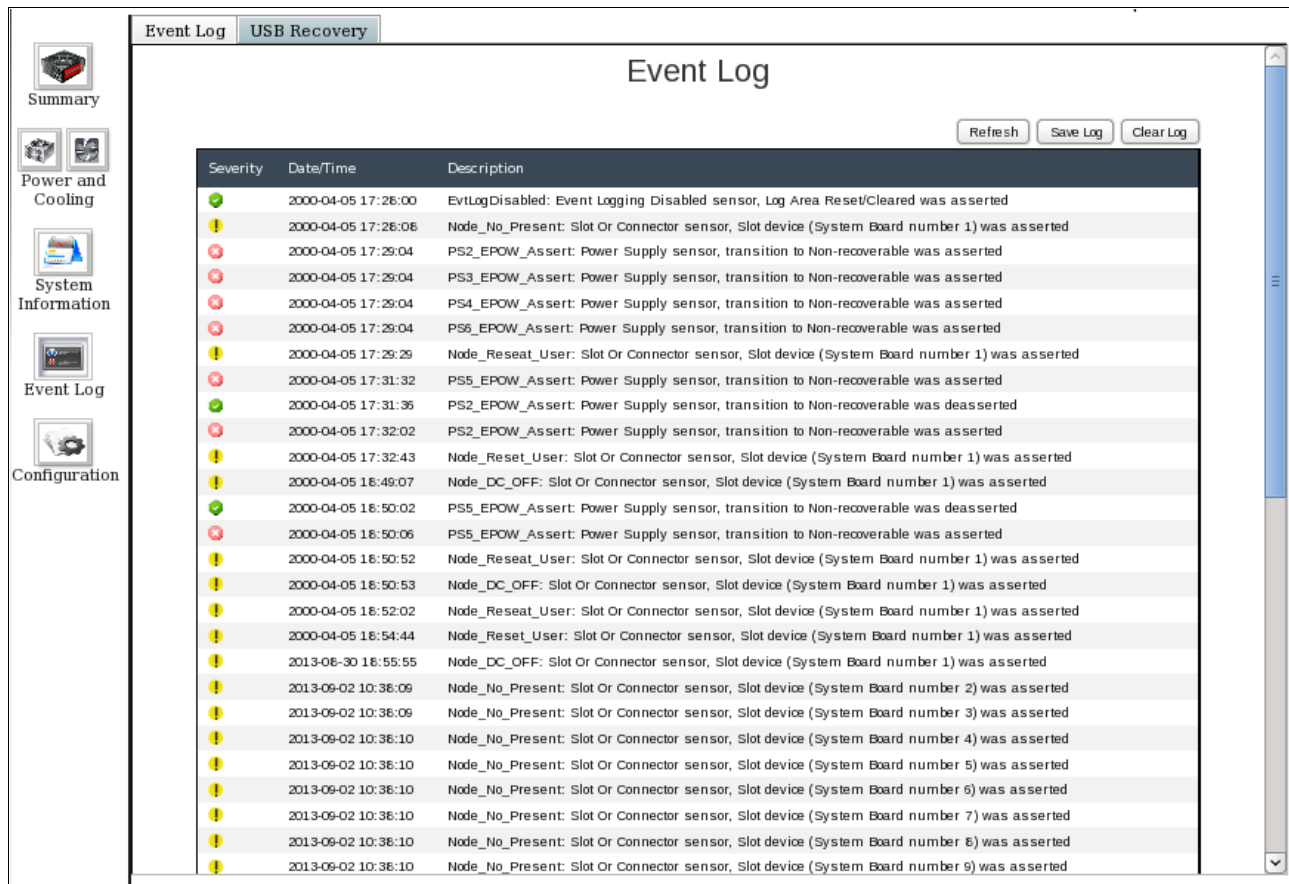


Figure 7-31 SEL from the FPC module

Full log: After the log fills up, you should clear it manually; otherwise, any other log entries cannot be received.

Back up and restore operations on the internal USB are automatically done by the FPC. Any change that is done through the web interface or the IPMI interface that is part of the following settings is saved in the internal USB:

- ▶ Selected power supply redundancy policy
- ▶ Oversubscription mode
- ▶ Power capping and power-saving values at chassis and node level
- ▶ Acoustic mode settings
- ▶ Power restore policy
- ▶ System event log (SEL)

All of these settings are volatile, so when FPC is rebooted, the FPC restores the settings from the internal USB automatically. The configuration settings that are related to network, SNMP, and so on, are non-volatile, so they remain in FPC memory between reboots.

The FPC web interface also includes manual backup and restore functions, as shown in Figure 7-28; however, because backup and restore tasks are automatic, these options are not needed.

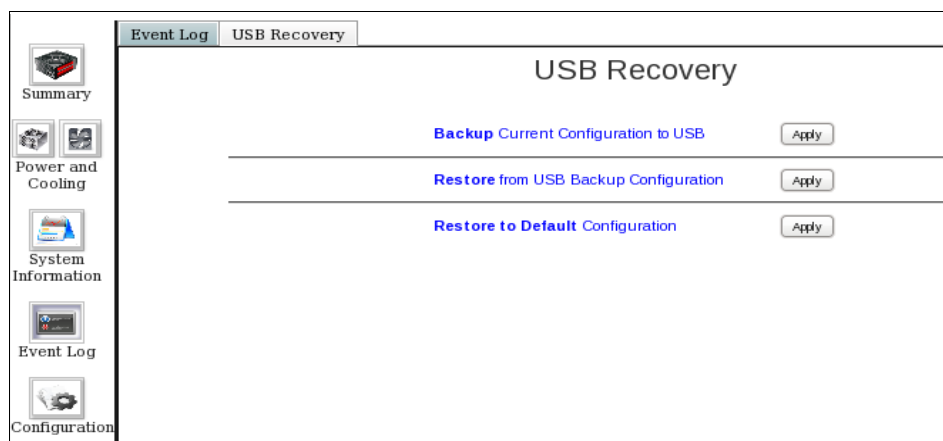


Figure 7-32 USB backup and recovery tab

Configuration

The Configuration function displays and configures the FPC module. All settings under the Configuration function are non-volatile, so they are kept between FPC reboots and are not saved to the internal USB key.

By using the Configuration function, the user can perform the following tasks by using the corresponding tabs:

- ▶ “Firmware Update tab”
- ▶ “SMTP tab” on page 137
- ▶ “SNMP tab” on page 138
- ▶ “Platform Filter Events tab” on page 138
- ▶ “Network Configuration tab” on page 138
- ▶ “Time Setting tab” on page 139
- ▶ “User Account tab” on page 139
- ▶ “Web Service tab” on page 140

Firmware Update tab

When a firmware upgrade is available, you use the web interface to perform the upgrade.

A firmware update is done in two phases by using the window that is shown in Figure 7-29. First, the user selects the wanted local firmware file that is uploaded and verified to be valid. Second, after the firmware is checked, a confirmation is requested. A table shows the actual firmware version, the new firmware version, and a preserve existing settings option that must be selected to keep the settings.

After the firmware update is performed, the FPC is rebooted.

The screenshot shows the 'Firmware Update' tab selected in the top navigation bar. The left sidebar contains icons for Summary, Power and Cooling, System Information, Event Log, and Configuration. The main content area is titled 'Firmware Update' and contains the following sections:

- Upload:** A text box for 'Firmware File Path' with a 'Choose File' button and the path 'ibm_fw_fpc...noarch.rpm'. An 'Upload' button is to the right.
- Firmware Image:** A table with four columns: 'Current Version', 'New Version', 'Preserve Settings', and 'Recover Primary Bank Firmware'.

Current Version	New Version	Preserve Settings	Recover Primary Bank Firmware
2.09	2.10	<input checked="" type="checkbox"/>	<input type="checkbox"/>
- Status Message:** 'Upload is completed. Please click 'Update' to proceed firmware update or click 'Cancel' to terminate the update. System will be rebooted after Update/Cancel process.'
- Buttons:** 'Update' and 'Cancel' buttons at the bottom.

Figure 7-33 Selection window for firmware update

SMTP tab

The FPC module allows the SMTP configuration to send the events to the destination email addresses and SMTP server, as shown in Figure 7-30. The Global Alerting Enable option at the Platform Event Filters (PEF) tab must be selected to enable SMTP traps and no filtering applied so that all the events are sent.

The screenshot shows the 'SMTP' tab selected in the top navigation bar. The left sidebar is the same as in Figure 7-33. The main content area is titled 'SMTP' and contains the following sections:

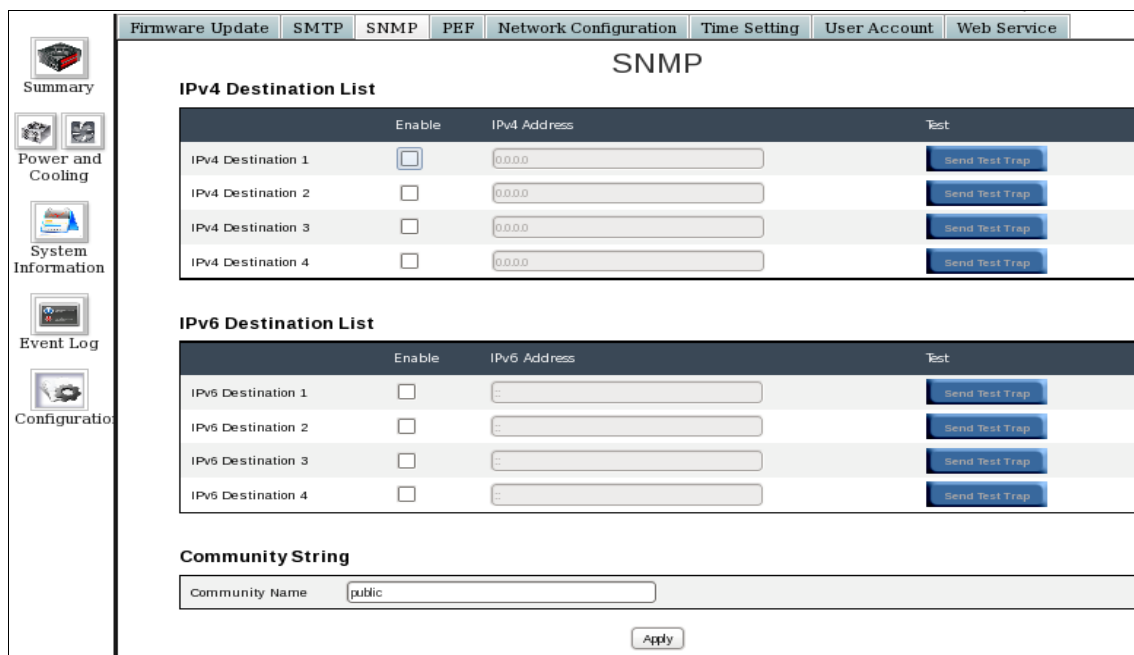
- Destination Email Addresses:** A table with four rows for 'Email Alert 1' through 'Email Alert 4'. Each row has an 'Enable' checkbox, a 'Destination Email Address' text box, an 'Email Description' text box (all containing 'MergePoint email ale'), and a 'Test' button (labeled 'Send Alert 1' through 'Send Alert 4').

Enable	Destination Email Address	Email Description	Test
<input type="checkbox"/>		MergePoint email ale	Send Alert 1
<input type="checkbox"/>		MergePoint email ale	Send Alert 2
<input type="checkbox"/>		MergePoint email ale	Send Alert 3
<input type="checkbox"/>		MergePoint email ale	Send Alert 4
- SMTP (email) Server Address:** A text box for 'SMTP IP Address' with the value '0.0.0.0'.
- SMTP Authentication:**
 - 'Enable' checkbox: ☐. A tooltip icon indicates 'Anonymous account will be used when authentication is disabled.'
 - 'Username' text box: empty.
 - 'Password' text box: empty.
 - 'STARTTLS Mode' dropdown: set to 'AUTO'.
 - 'SASL Mode' dropdown: set to 'AUTO'.
- Buttons:** An 'Apply' button at the bottom.

Figure 7-34 SMTP configuration tab

SNMP tab

The FPC module allows the SNMP configuration to send as SNMP traps the events that occur, as shown in Figure 7-31. The specific event types that are sent are selected at the PEF tab. The Global Alerting Enable option in the PEF tab must be selected so that the SNMP traps are enabled.



The image shows the SNMP configuration tab in a web interface. The top navigation bar includes tabs for Firmware Update, SMTP, SNMP, PEF, Network Configuration, Time Setting, User Account, and Web Service. The left sidebar contains icons for Summary, Power and Cooling, System Information, Event Log, and Configuration. The main content area is titled 'SNMP' and contains three sections: IPv4 Destination List, IPv6 Destination List, and Community String.

IPv4 Destination List

	Enable	IPv4 Address	Test
IPv4 Destination 1	<input checked="" type="checkbox"/>	0.0.0.0	<button>Send Test Trap</button>
IPv4 Destination 2	<input type="checkbox"/>	0.0.0.0	<button>Send Test Trap</button>
IPv4 Destination 3	<input type="checkbox"/>	0.0.0.0	<button>Send Test Trap</button>
IPv4 Destination 4	<input type="checkbox"/>	0.0.0.0	<button>Send Test Trap</button>

IPv6 Destination List

	Enable	IPv6 Address	Test
IPv6 Destination 1	<input type="checkbox"/>	-	<button>Send Test Trap</button>
IPv6 Destination 2	<input type="checkbox"/>	-	<button>Send Test Trap</button>
IPv6 Destination 3	<input type="checkbox"/>	-	<button>Send Test Trap</button>
IPv6 Destination 4	<input type="checkbox"/>	-	<button>Send Test Trap</button>

Community String

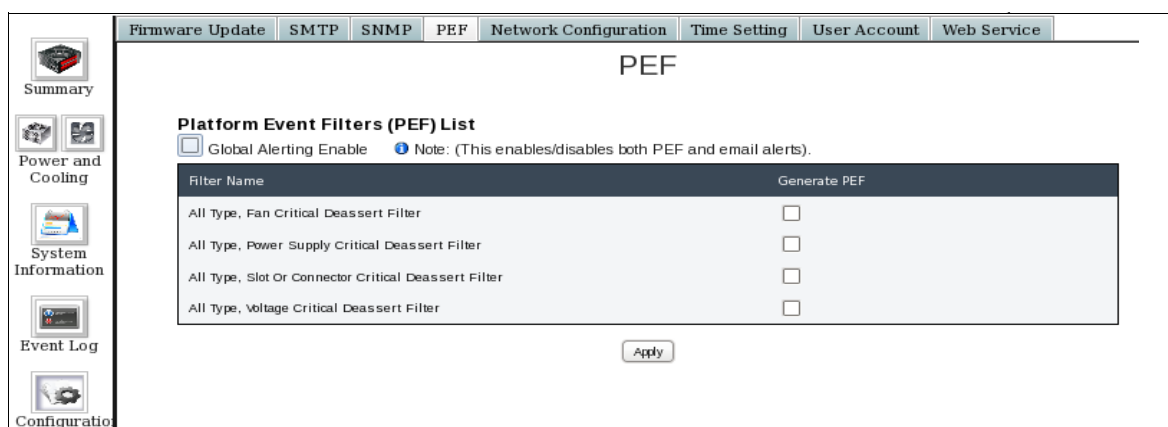
Community Name:

Apply

Figure 7-35 SNMP configuration tab

Platform Filter Events tab

In the PEF tab, you can configure the type of events that are sent as SNMP traps, as shown in Figure 7-32. You also can enable or disable SNMP and SMTP alerting by selecting the Global Alerting Enable option.



The image shows the Platform Event Filters (PEF) window in a web interface. The top navigation bar includes tabs for Firmware Update, SMTP, SNMP, PEF, Network Configuration, Time Setting, User Account, and Web Service. The left sidebar contains icons for Summary, Power and Cooling, System Information, Event Log, and Configuration. The main content area is titled 'PEF' and contains a section for Platform Event Filters (PEF) List.

Platform Event Filters (PEF) List

☐ Global Alerting Enable Note: (This enables/disables both PEF and email alerts).

Filter Name	Generate PEF
All Type, Fan Critical Deassert Filter	<input type="checkbox"/>
All Type, Power Supply Critical Deassert Filter	<input type="checkbox"/>
All Type, Slot Or Connector Critical Deassert Filter	<input type="checkbox"/>
All Type, Voltage Critical Deassert Filter	<input type="checkbox"/>

Apply

Figure 7-36 Platform Event Filters window

Network Configuration tab

In the Network Configuration tab, users can configure the network setting for the FPC module, as shown in Figure 7-33 on page 139. Hostname, static IP address, DHCP, and VLAN configuration can be set.

To access the specific network configuration settings windows, double-click the current network interface configuration.

Name	iF Enabled	IPv4 Enabled	IPv4 Address	IPv6 Enabled	IPv6 Address
eth0	Enabled	Enabled	192.168.0.100	Disabled	::0

Figure 7-37 Network configuration window.

Time Setting tab

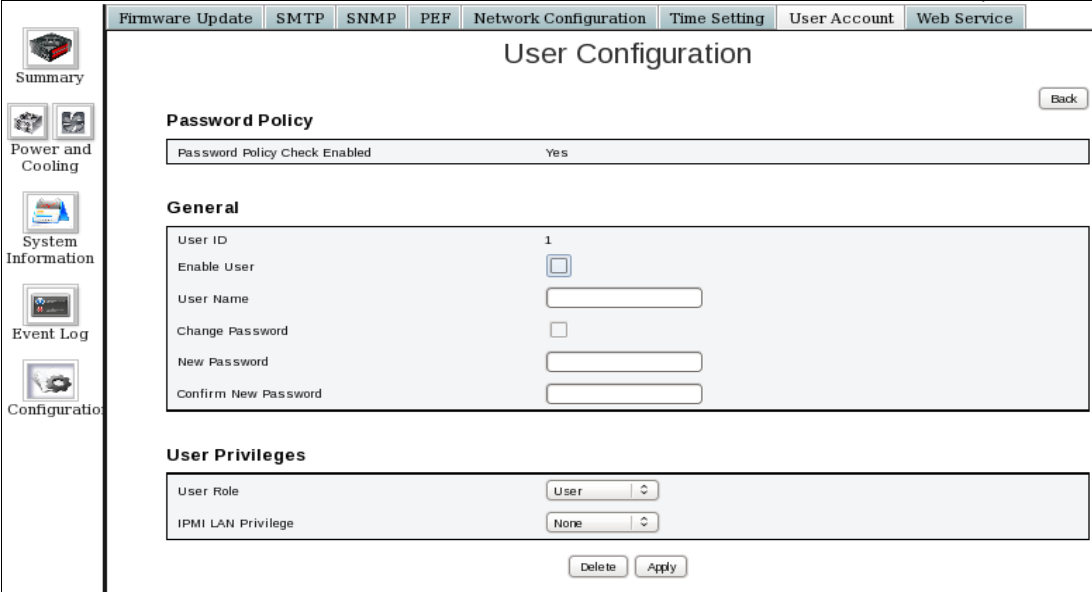
In the Time Setting tab, users can configure the date and time for the FPC module, as shown in Figure 7-34.

Figure 7-38 Date and time configuration window

User Account tab

In the User Account tab, users can add or remove users and assign one of the following user roles, as shown in Figure 7-35 on page 140:

- ▶ Administrator: Full access to all web pages and settings.
- ▶ Operator: Full access to all web pages and settings except the User Account page.
- ▶ User: Full access and settings to all pages except the Configuration function tab.

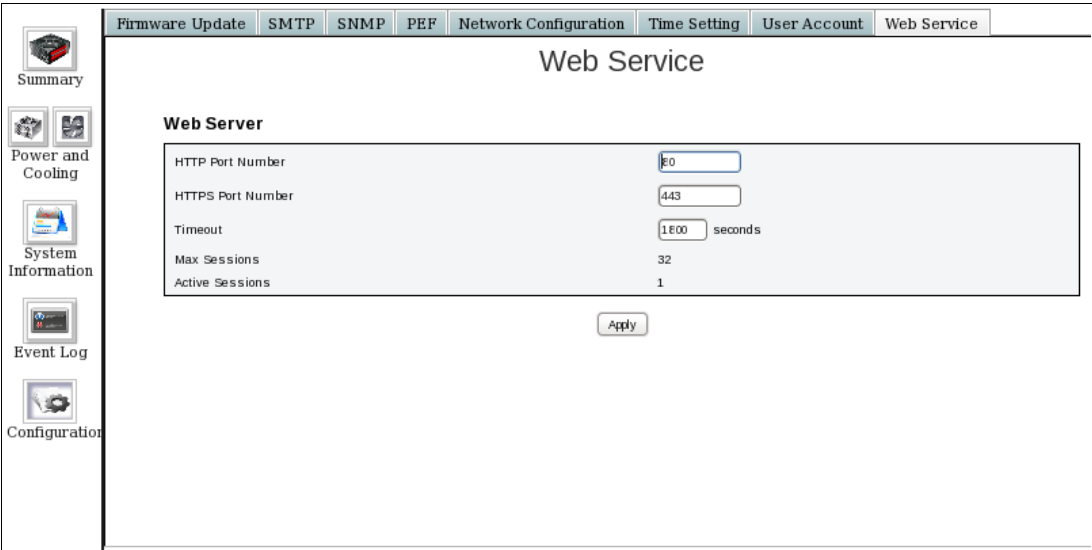


The screenshot shows the 'User Configuration' window in a web management interface. The left sidebar contains icons for Summary, Power and Cooling, System Information, Event Log, and Configuration. The top navigation bar includes tabs for Firmware Update, SMTP, SNMP, PEF, Network Configuration, Time Setting, User Account, and Web Service. The main content area is titled 'User Configuration' and includes a 'Back' button. It is divided into three sections: 'Password Policy' with a 'Password Policy Check Enabled' status set to 'Yes'; 'General' with fields for User ID (1), Enable User (checkbox), User Name (text input), Change Password (checkbox), New Password (text input), and Confirm New Password (text input); and 'User Privileges' with dropdown menus for User Role (User) and IPMI LAN Privilege (None). 'Delete' and 'Apply' buttons are at the bottom right.

Figure 7-39 User Configuration window

Web Service tab

User can configure the web interface ports for HTTP and HTTPS access in the Web Service tab, as shown in Figure 7-36.



The screenshot shows the 'Web Service' window in the same web management interface. The left sidebar and top navigation bar are identical to the previous figure. The main content area is titled 'Web Service' and includes a 'Web Server' section with a light blue background. This section contains five fields: HTTP Port Number (80), HTTPS Port Number (443), Timeout (1800 seconds), Max Sessions (32), and Active Sessions (1). An 'Apply' button is located at the bottom right of the 'Web Server' section.

Figure 7-40 Configuration of the HTTP/HTTPS ports for the web browser interface

7.2.2 FPC IPMI interface

The FPC module on the NeXtScale n1200 Enclosure supports IPMI over LAN access. The FPC complies with IPMI v2.0 standard and uses extensions and OEM IPMI commands to access FPC module-specific features. The IPMI interface can be used to develop wrappers with which users can remotely manage and configure multiple FPC modules at the same time.

By using the **ipmitool** command, you can manage and configure devices that support IPMI. The **ipmitool** command provides an easy CLI to start IPMI commands to a remote service processor through LAN. By using the tool, you also can send raw commands that are not part of the IPMI v2.0 definition but are vendor extensions to support certain specificities.

The FPC is compliant with IPMI v2.0, so default syntax and options can be used to access the commands that are part of the IPMI v2.0 definition. For example, to list the SEL, the following command can be used:

```
ipmitool -I lanplus -U USERID -P PASSWORD -H 192.168.0.100 sel list
```

OEM IPMI command extensions require the use of the raw interface **ipmitool** provides. The syntax for such interface has the following format:

```
ipmitool -I lanplus -U USERID -P PASSWORD \  
-H 192.168.0.100 raw <netfn> <cmd> [<byte1>] [<byte2>] ...
```

Output is raw hexadecimal data that provides the completion code of the operations and the resulting data. Input parameters and output values are described next. “Examples” on page 146 provides some examples of how IPMI commands are sent and how to process the output.

Note: Cluster management software provides command line tools to use this interface easier. However, some might not be implemented or you might want to integrate IBM NeXtScale System into your own custom monitoring infrastructure, so the IPMI commands are provided as reference.

Many settings that are available at the web browser interface are provided as IPMI commands. The following tables group the specific commands according to their functionality:

- ▶ Table 7-3 on page 142 lists Power supply unit (PSU) IPMI commands
- ▶ Table 7-4 on page 142 lists Power capping IPMI commands
- ▶ Table 7-5 on page 143 lists Power redundancy IPMI commands
- ▶ Table 7-6 on page 144 lists Acoustic modes IPMI commands
- ▶ Table 7-7 on page 144 lists Power restore policy IPMI commands
- ▶ Table 7-8 on page 144 lists Fan IPMI commands
- ▶ Table 7-9 on page 145 lists LED IPMI commands
- ▶ Table 7-10 on page 145 lists Node IPMI commands
- ▶ Table 7-11 on page 146 lists Miscellaneous IPMI commands
- ▶ “Examples” on page 146 shows the usage of IPMI interface with some examples

Table 7-4 Power supply unit (PSU) IPMI commands

Description	NetFn	CMD	Data
Get PSU Data	0x32	0x90	<p>Request Data:</p> <p>Byte 1 options:</p> <ul style="list-style-type: none"> ► 1: AC-IN ► 2: DC-OUT ► 3: PSU fan power <p>Response Data:</p> <p>(when AC-IN, DC-OUT)</p> <p>Byte 1: Completion code (0x00)</p> <p>Byte 2: Sum of MIN AC-IN /(DC-OUT) Least Significant Bit (LSB)</p> <p>Byte 3: Sum of MIN AC-IN /(DC-OUT) Most Significant Bit (MSB)</p> <p>Byte 4: Sum of average AC-IN /(DC-OUT) LSB</p> <p>Byte 5: Sum of average AC-IN/(DC-OUT) MSB</p> <p>Byte 6: Sum of MAX AC-IN /(DC-OUT) LSB</p> <p>Byte 7: Sum of MAX AC-IN /(DC-OUT) MSB</p> <p>(when Fan power)</p> <p>Byte 1: Completion code (0x00)</p> <p>Byte 2: Sum of FAN_Power LSB</p> <p>Byte 3: Sum of FAN_Power Byte 2</p> <p>Byte 4: Sum of FAN_Power MSB</p>
Get PSU Status	0x32	0x91	<p>Request Data:</p> <p>None</p> <p>Response Data:</p> <p>Byte 1: Completion code (0x00)</p> <p>Byte 2: PS_EPOW</p> <p>Byte 3: PS_THROTTLE</p> <p>Byte 4: PS_PRESENT</p> <p>Byte 5: PS_PWR_GOOD</p> <p>Byte 6: EPOW_OUT</p> <p>Byte 7: THROTTLE</p> <p>Each Byte is a bit mask where bit 0-5 = PSU1-6 (0: not trigger; 1: trigger)</p>

Table 7-5 Power capping IPMI commands

Description	NetFn	CMD	Data
Get power capping capacity	0x32	0x9d	<p>Request Data:</p> <p>Byte 1 options:</p> <ul style="list-style-type: none"> ► Node number: 0x1 to 0x0c for Node 1 - 12 ► Chassis: 0x0d <p>Response Data:</p> <p>Byte 1: Completion code (0x00) or out of range (0xC9)</p> <p>Byte 2: Min. capping value (LSB)</p> <p>Byte 3: Min. capping value (MSB)</p> <p>Byte 4: Max. capping value (LSB)</p> <p>Byte 5: Max. capping value (MSB)</p>
Set power capping value	0x32	0x9e	<p>Request Data:</p> <p>Byte 1 options:</p> <ul style="list-style-type: none"> ► Node number: 0x1 to 0x0c for Node 1 - 12 ► Chassis: 0x0d <p>Byte 2: Capping value LSB</p> <p>Byte 3: Capping value MSB</p> <p>Response Data:</p> <p>Byte 1: completion code (0x00) or out of range (0xC9) or cur not support (0xD5)</p>

Description	NetFn	CMD	Data
Set power-saving state	0x32	0x9f	Request Data: Byte 1 options: <ul style="list-style-type: none"> ▶ Node number: 0x1 to 0x0c for Node 1 - 12 ▶ Chassis: 0x0d Byte 2: Capping disable / enable Byte 3: Saving mode (0x00: disable; 0x01: Mode1; 0x02: Mode2; 0x03: Mode3) Response Data: Byte 1: Completion code (0x00) or out of range (0xC9)
Get power-saving state	0x32	0xa0	Request Data: Byte 1 options: <ul style="list-style-type: none"> ▶ Node number: 0x1 to 0x0c for Node 1 - 12 ▶ Chassis: 0x0d Response Data: Byte 1: Completion code (0x00) or out of range (0xC9) Byte 2: Capping disable / enable Byte 3: Capping value LSB Byte 4: Capping value MSB Byte 5: Saving mode

Table 7-6 Power redundancy IPMI commands

Description	NetFn	CMD	Data
Get PSU policy	0x32	0xa2	Request Data: None Response Data: Byte 1: Completion code (0x00) Byte 2: PSU Policy <ul style="list-style-type: none"> ▶ 0: No redundancy ▶ 1: N+1 ▶ 2: N+N Byte 3: Oversubscription mode (0: disable; 1: enable) Byte 4: Power bank LSB Byte 5: Power bank MSB
Set PSU policy	0x32	0xa3	Request Data: Byte 1: PSU Policy <ul style="list-style-type: none"> ▶ 0: No redundancy ▶ 1: N+1 ▶ 2: N+N Response Data: Byte 1: Completion code (0x00) or out of range (0xC9) or config not allowed (0x01) or bank lack (0x02)
Set Over Subscription mode	0x32	0x9c	Request Data: Byte 1: Over Subscription mode <ul style="list-style-type: none"> ▶ 0: Disable ▶ 1: Enable Response Data: Byte 1: Completion code (0x00) or cur not supported (0x5d) or param out of range (0xc9)

Table 7-7 Acoustic mode IPMI commands

Description	NetFn	CMD	Data
Set Acoustic mode	0x32	0x9b	Request Data: Byte 1: Acoustic mode (0x00: disable; 0x01: mode1 - 28%; 0x02; mode2 - 34%; 0x3 mode3 - 40%) Response Data: Byte 1: Completion code (0x00) or out of range (0xC9)

Table 7-8 Power restore policy IPMI commands

Description	NetFn	CMD	Data
Get Restore Policy	0x32	0xa9	Request Data: Byte 1: Node number LSB (bit mask) Byte 2: Node number LSB (bit mask) example: If setting 1,2 and 3 - Byte 1: 0x7 (0000 0111) Response Data: Byte 1: Completion code (0x00) or out of range (0xC9)
Set Restore Policy	0x32	0xaa	Request Data: None Response Data: Byte 1: Completion code (0x00) or out of range (0xC9) Byte 2: Node number LSB (bit mask) Byte 3: Node number LSB (bit mask)

Table 7-9 Fan IPMI commands

Description	NetFn	CMD	Data
Get PSU Fan status	0x32	0xa5	Request Data: Byte 1: PSU FAN number (0x01-0x06 FAN 1-6) Response Data: Byte 1: Fan speed LSB (rpm) Byte 2: Fan speed MSB (rpm) Byte 3: Fan speed (0-100%) Byte 4: Fan health <ul style="list-style-type: none"> ▶ 0: Not present ▶ 1: Abnormal ▶ 2: Normal

Table 7-10 LED IPMI commands

Description	NetFn	CMD	Data
Get Sys LED: Command to get FPC LED status	0x32	0x96	Request Data: None Response Data: Byte 1: Completion code (0x00) Byte 2: SysLocater LED Byte 3: CheckLog LED Possible values are 0: Off 1:On 2: Blink (SysLocater LED only)
Set Sys LED: Command to set FPC LED status	0x32	0x97	Request Data: Byte 1 options: ▶ 1: SysLocater LED ▶ 2: CheckLog LED Byte 2 options: ▶ 0: Disable ▶ 1: Enable ▶ 2: Blink (SysLocated LED only) Response Data: Byte 1: Completion code (0x00)

Table 7-11 Node IPMI commands

Description	NetFn	CMD	Data
Get Node Status	0x32	0xa7	Request Data: Byte 1: Node number: 0x1 to 0x0c for Node 1 - 12 Response Data: Byte 1: Completion code (0x00) or out of range (0xC9) Byte 2: Node Power State (0x00: Power OFF; 0x10: S3; 0x20: No permission fail; 0x40: Fault; 0x80: Power ON) Byte 3: Width Byte 4: Height Byte 5: Permission state (0x00 Not present; 0x01: Standby; 0x02: First permission fail; 0x03: Second permission fail; 0x04: Permission pass)
Reset/reseat Node	0x32	0xa4	Request Data: Byte 1: Node number: 0x1 to 0x0c for Node 1 - 12 Byte 2: Reset action: 1: reset, 2: reseat Response Data: Byte 1 – completion code (0x00) or cur not support (0x0xd5)
Show information about node size	0x32	0x99	Request Data: Byte 1: Node number: 0x1 to 0x0c for Node 1 - 12 Response Data: Byte 1: Completion code (0x00) or out of range (0xC9) Byte 2: Node Physical Width Byte 3: Node Physical Height Byte 4: Add-on Valid Byte 5: Add-on Width Byte 6: Add-on Height

Description	NetFn	CMD	Data
Show Node Power Consumption in watts	0x32	0x98	Request Data: Byte 1 options: <ul style="list-style-type: none"> Node number: 0x1 to 0x0c for Node 1 - 12 Chassis: 0x0d Response Data: Byte 1: Completion code (0x00) Byte 2: Power minimum (LSB) Byte 3: Power minimum (MSB) Byte 4: Power average (LSB) Byte 5: Power average (MSB) Byte 6: Power maximum (LSB) Byte 7: Power maximum (MSB)

Table 7-12 Miscellaneous IPMI commands

Description	NetFn	CMD	Data
Set Time	0x32	0xa1	Request Data: Byte 1: Year MSB (1970 - 2037) Byte 2: Year LSB (1970 - 2037) Byte 3: Month (0x01-0x12) Byte 4: Date (0x01-0x31) Byte 5: Hour (0x00-0x23) Byte 6: Minute (0x00-0x59) Byte 7: Second (0x00-0x59) Example: Year 2010 (byte1: 0x20; byte2: 0x10) Response Data: Byte 1: Completion code (0x00)
Get FPC Status	0x32	0xa8	Request Data: None Response Data: Byte 1: Completion code (0x00) Byte 2: FPC major version Byte 3: FPC minor version Byte 4: PSOC major version Byte 5: PSOC minor version Byte 6: Boot Flash number (0x1-0x2) Byte 7: Build major number Byte 8: Build minor number (ASCII value)

Examples

This section provides some examples of how to use the IPMI interface to obtain data. Depending on the command that is requested, the parameters and the output have a different format. For more information about specific command formats, see the following tables:

- ▶ Table 7-3 on page 142
- ▶ Table 7-4 on page 142
- ▶ Table 7-5 on page 143
- ▶ Table 7-6 on page 144
- ▶ Table 7-7 on page 144
- ▶ Table 7-8 on page 144
- ▶ Table 7-9 on page 145
- ▶ Table 7-10 on page 145
- ▶ Table 7-11

Get power consumption of a node

To get power consumption of node 1 (idle node), use the following command (see Table 7-10 on page 145 for the command syntax):

```
ipmitool -I lanplus -U USERID -P PASSWORD -H 192.168.0.100 raw 0x32 0x98 0x1
00 2c 00 2c 00 2e 00
```

The output string has the following meaning:

- ▶ Byte 1: Completion code 0x00
- ▶ Byte 2 and 3: Power minimum: 0x002C (44 W)
- ▶ Byte 4 and 5: Power average: 0x002C (44 W)
- ▶ Byte 6 and 7: Power maximum: 0x002E (46 W)

To get the power consumption of node 3, use the following command (see Table 7-10 on page 145 for the syntax):

```
ipmitool -I lanplus -U USERID -P PASSWORD -H 192.168.0.100 raw 0x32 0x98 0x3
00 93 00 94 00 97 00
```

The output string has the following meaning:

- ▶ Byte 1: Completion code 0x00
- ▶ Byte 2 and 3: Power minimum: 0x0093 (147 W)
- ▶ Byte 4 and 5: Power average: 0x0094 (148 W)
- ▶ Byte 6 and 7: Power maximum: 0x0097 (151 W)

Get status of a node

To get the status of node 1, use the following command (see Table 7-10 on page 145 for the syntax):

```
ipmitool -I lanplus -U USERID -P PASSWORD -H 192.168.0.100 raw 0x32 0xa7 0x1
00 80 01 01 04
```

The output string has the following meaning:

- ▶ Byte 1: Completion code 0x00
- ▶ Byte 2: Node power state 0x80 (Power ON)
- ▶ Byte 3: Width 0x01 (1U)
- ▶ Byte 4: Height 0x01 (1U)
- ▶ Byte 5: Permission state 0x04 (Permission pass)

Get power supply fan status

To get power supply fan status, use the following command (see Table 7-8 on page 144 for the syntax):

```
ipmitool -I lanplus -U USERID -P PASSWORD -H 192.168.0.100 raw 0x32 0xa5 0x1
b0 14 14 02
```

The output string has the following meaning:

- ▶ Byte 1 and 2: Fan speed 0x14B0 (5296 rpm)
- ▶ Byte 3: Fan speed% 0x14 (20%)
- ▶ Byte 4: Fan speed status 0x02 (normal)

Set acoustic mode

To set acoustic mode of the chassis to Mode 2, use the following command (see Table 7-6 on page 144 for the syntax):

```
ipmitool -I lanplus -U USERID -P PASSWORD -H 192.168.0.100 raw 0x32 0x9b 0x2
00
```

In the output string, byte 1 refers to completion code 0x0.

7.3 ServeRAID C100 drivers: nx360 M4

The ServeRAID C100 is an integrated SATA controller with software RAID capabilities. It is a cost-effective way to provide reliability, performance, and fault-tolerant disk subsystem management to help safeguard your valuable data and enhance availability.

IBM ServeRAID C100 RAID support must be enabled by pressing F1 at the setup menu.

RAID support for Windows and Linux only: No RAID support for VMware, Hyper-V, or Xen; for these operating systems, it can be used as a non-RAID SATA controller only.

By using the F1 setup menu, the MegaCLI command line utility, and the MegaRAID Storage Manager, a storage configuration must be created for the use of the software RAID capabilities. For more information about the setup and configuration instructions, see the ServeRAID C100 User's Guide, which is available at this website:

<http://ibm.com/support/entry/portal/docdisplay?ln docid=MIGR-5089055>

Operating System device drivers must be installed to use the ServeRAID C100 software RAID capabilities. The drivers are available at this website:

<http://ibm.com/support/entry/myportal/docdisplay?ln docid=MIGR-5089068>

You also can find the latest device drives for the different operating systems that support ServeRAID C100 controller and download links to the configuration tools, MegaCLI, and MegaRAID Storage Manager, with which you can create a storage configuration under the respective operating system.

7.4 Integrated SATA controller: nx360 M5

The NeXtScale nx360 M5 features an onboard SATA controller that is integrated in to the Intel C612 chipset. It supports one 3.5 inch simple-swap SATA or near line (NL) SATA drive, two 2.5 inch simple-swap NL SATA drives, or four 1.8-inch SATA solid-state drives (SSDs). Two 2.5-inch simple swap NL SATA drives or four 1.8-inch SATA SSDs can also be used with a RAID controller or SAS HBA that is installed in the internal RAID adapter riser slot.

Two 2.5-inch simple swap SAS drives or two 2.5-inch hot swap drives that are installed in the front drive bays require a RAID controller or SAS HBA that is installed in the internal RAID adapter riser slot.

7.5 VMware vSphere Hypervisor

The NeXtScale compute nodes support VMware vSphere Hypervisor (ESXi) that is installed on a USB memory key. The server provides the option of adding a blank USB memory key for the installation of the embedded VMware ESXi.

The VMware ESXi embedded hypervisor software is a virtualization platform with which multiple operating systems can be run on a host system at the same time.

Lenovo provides different versions of VMware ESXi customized for IBM hardware that can be downloaded from this website:

http://shop.lenovo.com/us/en/systems/solutions/alliances/vmware/#tab-vmware_vsphere_esxi

For more information about installation instructions, see *vSphere Installation and Setup Guide*, which is provided as part of the downloaded image.

7.6 eXtreme Cloud Administration Toolkit

The eXtreme Cluster Administration Toolkit (xCAT) 2 is an Open Source Initiative that was developed by IBM to support the deployment of large high-performance computing (HPC) clusters that are based on various hardware platforms. xCAT 2 is not an evolution of the earlier xCAT 1. Instead, it is a complete code write that combines the best practices of Cluster Systems Management (CSM) and xCAT 1. xCAT 2 is open to the general HPC community under the Eclipse License to help support and enhance the product in the future.

xCAT provides a scalable distributed computing management and provisioning tool that provides a unified interface for hardware control, discovery, remote control management, and operating system diskfull and diskless deployment.

xCAT 2 uses only scripts, which makes the code portable and modular in nature and allows for the easy inclusion of more functions and plug-ins.

The xCAT architecture includes the following main features:

- ▶ Client/server architecture
Clients can run on any Perl-compliant system (including Windows). All communications are SSL encrypted.
- ▶ Role-based administration
Different users can be assigned various administrative roles for different resources.
- ▶ Stateful, stateless, and iSCSI nodes provisioning support
Stateless nodes can be RAM-root, compressed RAM-root, or stacked NFS-root. Linux software initiator iSCSI support for Red Hat Enterprise Linux (RHEL) and SUSE Linux Enterprise Server (SLES) is included.

Systems without hardware-based initiators also can be installed and booted by using iSCSI.

► Scalability

xCAT 2 can scale to 100,000 and more nodes with xCAT's Hierarchical Management Cloud. A single management node can have any number of stateless service nodes to increase the provisioning throughput and management of the largest clusters. All cluster services, such as, LDAP, DNS, DHCP, NTP, and Syslog are configured to use the Hierarchical Management Cloud. Outbound cluster management commands (for example, **rpower**, **xdsh**, and **xdcp**) use this hierarchy for scalable systems management. An example of such a hierarchy is shown in Figure 7-37.

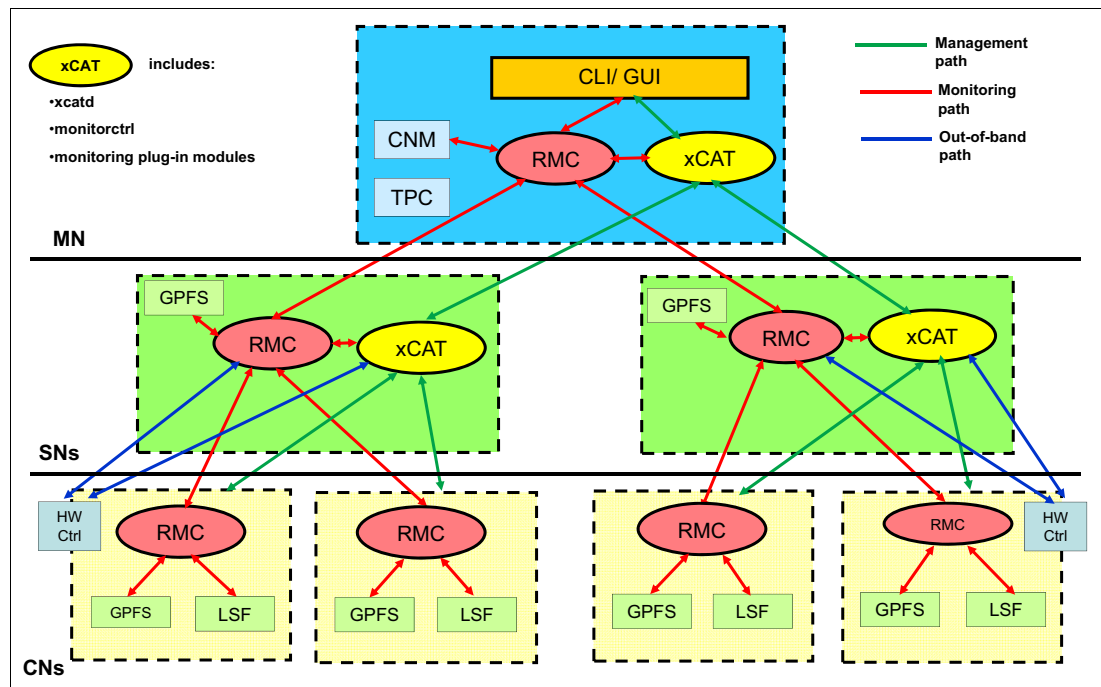


Figure 7-41 xCAT hierarchical cluster management

► Automatic discovery

This feature includes single power button press, physical location-based discovery, and configuration capability. Although this feature is mostly hardware-dependent, xCAT 2 was developed to ease integration for new hardware. The plug-in software architecture provides an easy development mechanism for the new hardware.

► Plug in architecture for compartmental development

By using this feature, you can add your own xCAT functionality to do whatever you want. New plug-ins extend the xCAT vocabulary that is available to xCAT clients.

► Notification infrastructure

By using this feature, you can watch for xCAT DB table changes via the notification infrastructure.

► SNMP monitoring

Default monitoring uses SNMP trap handlers to handle all SNMP traps.

► Flexible monitoring infrastructure

You can easily integrate third-party vendor monitoring software into the xCAT cluster. Currently, the following plug-ins are provided with xCAT:

- SNMP
- RMC (RSCT)

- Ganglia
- Performance Copilot
- ▶ Centralized console and system logs

xCAT provides console access to managed nodes and centralized logging.

xCAT 2 evolved to support several operating systems, including AIX, and the many derivatives of Linux, including SLES, openSUSE, RHEL, CentOS, and Fedora Core. xCAT also can provision Windows 2008 through imaging and virtual machine (VMware, Xen, KVM) images to hosted systems. Because xCAT is an open source project, other operating systems can be added to the supported list in the future.

With AIX and Linux, xCAT supports traditional local disk, SAN disk, and stateful diskless, which provisions via native deployment methods. Also, support is provided for stateless diskless nodes including ramfs root, compressed ramfs root, and NFS root with ramfs overlay support (Linux and AIX) and stateful diskless that uses iSCSI (Linux).

The xCAT manager works with the relevant hardware control units within the nodes, BladeCenter, and HMC to instruct these units to perform a number of hardware functions or gather information about the hosts.

The following hardware control features are included:

- ▶ Power control (power on, off, cycle, and current state)
- ▶ Event logs
- ▶ Boot device control (full boot sequence on IBM System BladeCenter, next boot device on other systems)
- ▶ Sensor readings (temperature, fan speed, voltage, current, and fault indicators as supported by systems)
- ▶ Node MAC address gathering
- ▶ LED status and modification (ability to identify LEDs on all systems, and diagnostic LEDs on select IBM rack-mount servers)
- ▶ Serial-over-LAN (SOL): Use or redirect input and output
- ▶ Service processor configuration
- ▶ Hardware control point discovery by using Service Location Protocol (SLP): BladeCenter Advanced Management Module (AMM), IBM Power Systems HMC, and Flexible Service Processor (FSP).

Supported hardware

The following hardware is supported and tested by xCAT:

- ▶ IBM BladeCenter with AMM
- ▶ IBM System x
- ▶ IBM Power Systems (including the HMC)
- ▶ IBM iDataPlex
- ▶ IBM Flex System
- ▶ IBM NeXtscale System
- ▶ Machines that are based on the IPMI

Because only IBM hardware is available for testing, this hardware is the only hardware that is supported. However, other vendors' hardware can be managed with xCAT. For more information and the latest list of supported hardware, see this website:

<http://xcat.sourceforge.net/>

Note: Although xCAT is an Open Source Initiative that is provided under the Eclipse license and freely available, official IBM support can be contacted.

Abbreviations and acronyms

AC	alternating current	DIMM	dual inline memory module
ACPI	advanced control and power interface	DNS	Domain Name System
AMM	Advanced Management Module	DOS	disk operating system
ASHRAE	American Society of Heating, Refrigerating, and Air-Conditioning Engineers	DP	dual processor
ASR	automatic server restart	DPC	deferred procedure call
ASU	Advanced Settings Utility	DRAM	dynamic random access memory
AVX	Advanced Vector Extensions	DSA	Dynamic System Analysis
BIOS	basic input output system	DVD	Digital Video Disc
BMC	Baseboard Management Controller	DWC	direct water cooling
BSMI	Bureau of Standards, Metrology and Inspection	ECC	error checking and correcting
CB	Certification Body	EDR	Enhanced Data Rate
CCC	China Compulsory Certificate	EMI	Electromagnetic interference
CD	compact disk	EPOW	Early Power Off Warning
CD-ROM	compact disc read only memory	FAN	Fabric Address Notification
CDU	chiller distribution unit	FC	Fibre Channel
CE	Conformité Européenne	FCC	Federal Communications Commission
CFF	common form factor	FCP	Flow Control Packet
CFM	cubic feet per minute	FDR	fourteen data rate
CIM	Common Information Model	FHHL	full-height half-length
CISPR	International Special Committee on Radio Interference	FLOPS	floating-point operations per second
CLI	command-line interface	FMA	fused multiply add
CMD	command	FOD	features on demand
CMOS	complementary metal oxide semiconductor	FP	floating point
CNA	Converged Network Adapter	FPC	Fan and Power Controller
COD	configuration on disk	FSP	Flexible Service Processor
CPU	central processing unit	GB	gigabyte
CRC	cyclic redundancy check	GFLOPS	giga floating-point operations per second
CSA	Canadian Standards Association	GOST	gosudarstvennyy standart (state standard)
CSM	Cluster Systems Management	GPU	Graphics Processing Unit
CTO	configure-to-order	GT	Gigatransfers
DB	database	GUI	graphical user interface
DC	domain controller	HBA	host bus adapter
DDF	Disk Data Format	HCA	host channel adapter
DDR	Double Data Rate	HD	high definition
DHCP	Dynamic Host Configuration Protocol	HDD	hard disk drive
		HMC	Hardware Management Console
		HPC	high performance computing

HPL	High-Performance Linpack	NFS	network file system
HS	hot swap	NIC	network interface card
HTTP	Hypertext Transfer Protocol	NL	nearline
HW	hardware	NTP	Network Time Protocol
I/O	input/output	NUMA	Non-Uniform Memory Access
I/OAT	I/O Acceleration Technology	NVGRE	Network Virtualization using Generic Routing Encapsulation
IB	InfiniBand	NVIDIA	
IBM	International Business Machines	OEM	other equipment manufacturer
ID	identifier	OFED	OpenFabrics Enterprise Distribution
IEC	International Electrotechnical Commission	OS	operating system
IEEE	Institute of Electrical and Electronics Engineers	OVS	oversubscription
IMM	integrated management module	PCH	Platform Controller Hub
IOPS	I/O operations per second	PCI	Peripheral Component Interconnect
IP	Internet Protocol	PCI-E	PCI Express
IPMI	Intelligent Platform Management Interface	PDU	power distribution unit
ISO	International Organization for Standards	PE	Preinstallation Environment
IT	information technology	PEF	platform event filtering
JBOD	just a bunch of disks	PET	Platform Event Trap
KB	kilobyte	PF	power factor
KVM	keyboard video mouse	PFA	Predictive Failure Analysis
LAN	local area network	PN	part number
LDAP	Lightweight Directory Access Protocol	PSOC	Programmable System-on-Chip
LED	light emitting diode	PSU	power supply unit
LFF	large form factor	PXE	Preboot eXecution Environment
LLC	last level cache	QDR	quad data rate
LOM	LAN on motherboard	QPI	QuickPath Interconnect
LP	low profile	QR Code	Quick Response Code
LRDIMM	load-reduced dual inline memory module	RAID	redundant array of independent disks
LSB	Least Significant Bit	RAS	remote access services; row address strobe
LSO	large segment offload	RDHX	Rear Door Heat eXchanger
MAC	media access control	RDIMM	registered DIMM
MB	megabyte	RDMA	Remote Direct Memory Access
MLC	multi-level cell	RDS	Reliable Datagram Sockets
MPI	Message Passing Interface	RHEL	Red Hat Enterprise Linux
MSB	Most Significant Bit	RMC	Resource Monitoring and Control
MSI	Message Signaled Interrupt	ROC	RAID-on-card
MTM	machine type model	ROM	read-only memory
MTU	maximum transmission unit	RPM	revolutions per minute
NC-SI	Network Controller-Sideband Interface	RSCT	Reliable Scalable Cluster Technology
		RSS	Receive-side scaling

RX	receive	TUV-GS	Technischer Überwachungs-Verein Geprüfte Sicherheit (TUV tested safety)
SAN	storage area network	TX	transmit
SAS	Serial Attached SCSI	UDIMM	Unbuffered DIMM
SATA	Serial ATA	UDP	user datagram protocol
SCTP	Stream Control Transmission Protocol	UEFI	Unified Extensible Firmware Interface
SD	sales and distribution	UPS	uninterruptible power supply
SDDC	Single Device Data Correction	URL	Uniform Resource Locator
SDR	Single Data Rate	USB	universal serial bus
SED	self-encrypting drive	UXSP	UpdateXpress System Packs™
SEL	System Event Log	VCCI	Voluntary Control Council for Interference
SFF	Small Form Factor	VFA	Virtual Fabric Adapter
SFP	small form-factor pluggable	VGA	video graphics array
SGMII	Serial Gigabit Media-independent Interface	VLAN	virtual LAN
SHMEM	Symmetric Hierarchical Memory	VM	virtual machine
SIG	special interest group	VPD	vital product data
SKU	stock keeping unit	VPI	Virtual Protocol Interconnect
SLES	SUSE Linux Enterprise Server	VXLAN	Virtual Extensible LAN
SLP	Service Location Protocol	WCT	Water Cool Technology
SMB	server message block	XML	Extensible Markup Language
SMP	symmetric multiprocessing		
SMTP	simple mail transfer protocol		
SNMP	Simple Network Management Protocol		
SOL	Serial over LAN		
SR-IOV	single root I/O virtualization		
SRP	Storage RDMA Protocol		
SS	simple swap		
SSD	solid state drive		
SSE	Streaming SIMD Extensions		
SSH	Secure Shell		
SSL	Secure Sockets Layer		
SSP	Serial SCSI Protocol		
STP	Spanning Tree Protocol		
TB	terabyte		
TCO	total cost of ownership		
TCP	Transmission Control Protocol		
TCP/IP	Transmission Control Protocol/Internet Protocol		
TDP	thermal design power		
TOE	TCP offload engine		
TPM	Trusted Platform Module		
TSS	Trusted Computing Group Software Stack		

Related publications

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

Lenovo Press publications

For more information, see the following Lenovo Press publications:

- ▶ *NeXtScale nx360 M5 Product Guide*, TIPS1195:
<http://lenovopress.com/tips1195>
- ▶ *NeXtScale nx360 M4 Product Guide*, TIPS1051:
<http://lenovopress.com/tips1051>
- ▶ *xREF: x86 Server Reference*:
<http://lenovopress.com/xref>
- ▶ *Lenovo NeXtScale System Water Cool Technology Planning and Implementation Guide*, SG24-8276:
<http://lenovopress.com/sg248276>
- ▶ *NeXtScale System M5 with Water Cool Technology Product Guide*, TIPS1241:
<http://lenovopress.com/tips1241>

Product publications

The following product publications are relevant as further information sources:

- ▶ *NeXtScale nx360 M5 Installation and Service Guide*:
<http://ibm.com/support/entry/portal/docdisplay?ln docid=MIGR-5096282>
- ▶ *NeXtScale nx360 M4 Installation and Service Guide*:
<http://ibm.com/support/entry/portal/docdisplay?ln docid=MIGR-5093697>
- ▶ *NeXtScale n1200 Enclosure Installation and Service Guide*:
<http://ibm.com/support/entry/portal/docdisplay?ln docid=MIGR-5093698>
- ▶ *42U 1100 mm Enterprise V2 Dynamic Rack and Dynamic Expansion Rack Installation Guide*:
<http://ibm.com/support/entry/portal/docdisplay?ln docid=MIGR-5089535>
- ▶ *Rear Door Heat eXchanger V2 Type 1756 Installation and Maintenance Guide*:
<http://ibm.com/support/entry/portal/docdisplay?ln docid=MIGR-5089575>

Online resources

The following websites are relevant as further information sources:

- ▶ NeXtScale System home page:
<http://shop.lenovo.com/us/en/systems/servers/high-density/nextscale-m5/>
- ▶ NeXtScale System Power Requirements Guide:
<http://ibm.com/support/entry/portal/docdisplay?lndocid=LNVO-POWINF>
- ▶ Lenovo Power Configurator:
<http://ibm.com/support/entry/portal/docdisplay?lndocid=LNVO-PWRCONF>
- ▶ Lenovo Press Product Guides for System x servers and options:
<http://lenovopress.com/systemx>
- ▶ Configuration and Option Guide:
<http://www.ibm.com/systems/xbc/cog/>
- ▶ System x Support Portal:
<http://ibm.com/support/entry/portal/>

Lenovo[™]

Lenovo NeXtScale System Planning and Implementation Guide



(0.5" spine)
0.475" <-> 0.873"
250 <-> 459 pages



Lenovo NeXtScale System Planning and Implementation Guide

Introduces the high density x86 solution for scale-out environments

Covers the air-cooled NeXtScale System offerings: nx360 M5 & M4 nodes, and n1200 enclosure

Addresses power, cooling, racking, and management

Provides the information you need for a successful implementation

Lenovo NeXtScale System is a dense computing offering based on our experience with iDataPlex and Flex System and with a tight focus on emerging and future client requirements. The NeXtScale n1200 Enclosure and NeXtScale nx360 M5 Compute Node are designed to optimize density and performance within typical data center infrastructure limits.

The 6U NeXtScale n1200 Enclosure fits in a standard 19-inch rack and up to 12 compute nodes can be installed into the enclosure. With more computing power per watt and the latest Intel Xeon processors, you can reduce costs while maintaining speed and availability.

This Lenovo Press publication is for customers who want to understand and implement a NeXtScale System solution. It introduces the offering and the innovations in its design, outlines its benefits, and positions it with other x86 servers. The book provides details about NeXtScale System components and the supported options. It also provides rack and power planning considerations and describes the ways that you can manage the system.

This book describes the air-cooled NeXtScale System offerings. For planning and implementation information about the water-cooled offering, NeXtScale System WCT, see the Lenovo Press publication *Lenovo NeXtScale System Water Cool Technology Planning and Implementation Guide*.



**BUILDING
TECHNICAL
INFORMATION
BASED ON
PRACTICAL
EXPERIENCE**

At Lenovo Press, we bring together experts to produce technical publications around topics of importance to you, providing information and best practices for using Lenovo products and solutions to solve IT challenges.