



# Sun™ InfiniBand Dual Port 4x DDR PCIe® Fabric Expansion Module User's Guide

---

Sun Microsystems, Inc.  
[www.sun.com](http://www.sun.com)

Part No. 820-7293-10  
April 2009, Revision A

Submit comments about this document at: <http://www.sun.com/documentation/feedback>

Copyright 2009 Sun Microsystems, Inc., 4150 Network Circle, Santa Clara, California 95054, U.S.A. All rights reserved.

Sun Microsystems, Inc. has intellectual property rights relating to technology that is described in this document. In particular, and without limitation, these intellectual property rights may include one or more of the U.S. patents listed at <http://www.sun.com/patents> and one or more additional patents or pending patent applications in the U.S. and in other countries.

This document and the product to which it pertains are distributed under licenses restricting their use, copying, distribution, and decompilation. No part of the product or of this document may be reproduced in any form by any means without prior written authorization of Sun and its licensors, if any.

Third-party software, including font technology, is copyrighted and licensed from Sun suppliers.

Parts of the product may be derived from Berkeley BSD systems, licensed from the University of California. UNIX is a registered trademark in the U.S. and in other countries, exclusively licensed through X/Open Company, Ltd.

Sun, Sun Microsystems, the Sun logo, Java, Sun Blade, docs.sun.com, and Solaris are trademarks or registered trademarks of Sun Microsystems, Inc., or its subsidiaries, in the U.S. and in other countries.

The OPEN LOOK and Sun™ Graphical User Interface was developed by Sun Microsystems, Inc. for its users and licensees. Sun acknowledges the pioneering efforts of Xerox in researching and developing the concept of visual or graphical user interfaces for the computer industry. Sun holds a non-exclusive license from Xerox to the Xerox Graphical User Interface, which license also covers Sun's licensees who implement OPEN LOOK GUIs and otherwise comply with Sun's written license agreements.

PCIe is a registered trademark of PCI-SIG. PCI Express is a registered trademark of PCI-SIG.

U.S. Government Rights—Commercial use. Government users are subject to the Sun Microsystems, Inc. standard license agreement and applicable provisions of the FAR and its supplements.

DOCUMENTATION IS PROVIDED "AS IS" AND ALL EXPRESS OR IMPLIED CONDITIONS, REPRESENTATIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT, ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT SUCH DISCLAIMERS ARE HELD TO BE LEGALLY INVALID.

---

Copyright 2009 Sun Microsystems, Inc., 4150 Network Circle, Santa Clara, Californie 95054, États-Unis. Tous droits réservés.

Sun Microsystems, Inc. possède les droits de propriété intellectuelle relatifs à la technologie décrite dans ce document. En particulier, et sans limitation, ces droits de propriété intellectuelle peuvent inclure un ou plusieurs des brevets américains listés sur le site <http://www.sun.com/patents>, un ou les plusieurs brevets supplémentaires ainsi que les demandes de brevet en attente aux les États-Unis et dans d'autres pays.

Ce document et le produit auquel il se rapporte sont protégés par un copyright et distribués sous licences, celles-ci en restreignent l'utilisation, la copie, la distribution, et la décompilation. Aucune partie de ce produit ou document ne peut être reproduite sous aucune forme, par quelque moyen que ce soit, sans l'autorisation préalable et écrite de Sun et de ses bailleurs de licence, s'il y en a.

Tout logiciel tiers, sa technologie relative aux polices de caractères, comprise, est protégé par un copyright et licencié par des fournisseurs de Sun.

Des parties de ce produit peuvent dériver des systèmes Berkeley BSD licenciés par l'Université de Californie. UNIX est une marque déposée aux États-Unis et dans d'autres pays, licenciée exclusivement par X/Open Company, Ltd.

Sun, Sun Microsystems, le logo Sun, Java, Sun Blade, docs.sun.com, et Solaris sont des marques de fabrique ou des marques déposées de Sun Microsystems, Inc., ou ses filiales, aux États-Unis et dans d'autres pays.

L'interface utilisateur graphique OPEN LOOK et Sun™ a été développée par Sun Microsystems, Inc. pour ses utilisateurs et licenciés. Sun reconnaît les efforts de pionniers de Xerox dans la recherche et le développement du concept des interfaces utilisateur visuelles ou graphiques pour l'industrie informatique. Sun détient une licence non exclusive de Xerox sur l'interface utilisateur graphique Xerox, cette licence couvrant également les licenciés de Sun implémentant les interfaces utilisateur graphiques OPEN LOOK et se conforment en outre aux licences écrites de Sun.

PCIe est une marque déposée de PCI-SIG. PCI Express est une marque déposée de PCI-SIG.

LA DOCUMENTATION EST FOURNIE "EN L'ÉTAT" ET TOUTES AUTRES CONDITIONS, DÉCLARATIONS ET GARANTIES EXPRESSES OU TACITES SONT FORMELLEMENT EXCLUES DANS LA LIMITE DE LA LOI APPLICABLE, Y COMPRIS NOTAMMENT TOUTE GARANTIE IMPLICITE RELATIVE À LA QUALITÉ MARCHANDE, À L'APTITUDE À UNE UTILISATION PARTICULIÈRE OU À L'ABSENCE DE CONTREFAÇON.



# Contents

---

## **Preface v**

### **1. Product Overview 1**

Product Description 1

Hardware and Software Requirements 2

Dual Port Connectivity 3

Interfaces 3

InfiniBand Interface 3

PCI Express Interface 3

Memory 3

System Memory 4

Regulatory Compliance 4

### **2. Installing the InfiniBand Fabric Expansion Module 5**

Installing the IB-FEM 5

▼ To Install the IB-FEM 5

### **3. InfiniBand Software on Linux 7**

InfiniBand Overview 7

InfiniBand Software for Linux 8

OpenFabrics Enterprise Distribution for Linux 10

Installing the InfiniBand Drivers on Linux	11
▼ To Install IB Drivers From the Linux Distribution Source	11
▼ To Install the OFED Package	13
▼ To Verify Driver Installation on Linux	18
<b>4. Internet Protocol Over InfiniBand</b>	<b>21</b>
Configuring IPoIB on Linux	21
▼ To Install IPoIB Driver	22
▼ To Change IPoIB Configuration Without Rebooting	23
▼ To Change IB Startup Behavior on Linux	24
<b>A. Specifications</b>	<b>27</b>
IB-FEM Specifications	27
<b>Index</b>	<b>29</b>

# Preface

---

This guide provides an overview, installation instructions and specifications of the Sun™ InfiniBand Dual Port 4x DDR PCIe® Fabric Expansion Module.

The instructions in this guide are designed for system administrators with experience installing network hardware and software.

---

## Related Documentation

Title	URL
Sun Blade Server Module Documentation	<a href="http://docs.sun.com/app/docs/prod/blade.srvr">http://docs.sun.com/app/docs/prod/blade.srvr</a>
Sun Blade 6048 Modular System Documentation	<a href="http://docs.sun.com/app/prod/blade.6048">http://docs.sun.com/app/prod/blade.6048</a>
<i>Sun Blade 6048 InfiniBand QDR Switched Network Express Module User's Guide (820-6705)</i>	<a href="http://docs.sun.com/app/prod/blade.6048">http://docs.sun.com/app/prod/blade.6048</a>
<i>Sun Blade 6048 InfiniBand QDR Switched Network Express Module Product Notes (820-6706)</i>	<a href="http://docs.sun.com/app/prod/blade.6048">http://docs.sun.com/app/prod/blade.6048</a>

---

# Documentation, Support, and Training

Sun Function	URL
Documentation	<a href="http://docs.sun.com">http://docs.sun.com</a>
Support	<a href="http://www.sun.com/support/">http://www.sun.com/support/</a>
Training	<a href="http://www.sun.com/training/">http://www.sun.com/training/</a>

---

---

## Third-Party Web Sites

Sun is not responsible for the availability of third-party web sites mentioned in this document. Sun does not endorse and is not responsible or liable for any content, advertising, products, or other materials that are available on or through such sites or resources. Sun will not be responsible or liable for any actual or alleged damage or loss caused by or in connection with the use of or reliance on any such content, goods, or services that are available on or through such sites or resources.

PCI Express® specifications can be found at this web site:

<http://www.pcisig.com>

Mellanox Technologies HCA information can be found at this web site:

<http://www.mellanox.com>

---

## Sun Welcomes Your Comments

Sun is interested in improving its documentation and welcomes your comments and suggestions. You can submit your comments by going to:

<http://www.sun.com/documentation/feedback>

Please include the title and part number of your document with your feedback:

*Sun InfiniBand Dual Port 4x DDR PCIe Fabric Expansion Module User's Guide*, part number 820-7293-10.

# Product Overview

---

This chapter provides an overview of the Sun™ InfiniBand Dual Port 4x DDR PCIe® Fabric Expansion Module (IB-FEM).

This chapter includes the following sections:

- [“Product Description” on page 1](#)
- [“Hardware and Software Requirements” on page 2](#)
- [“Interfaces” on page 3](#)
- [“Memory” on page 3](#)
- [“Regulatory Compliance” on page 4](#)

---

## Product Description

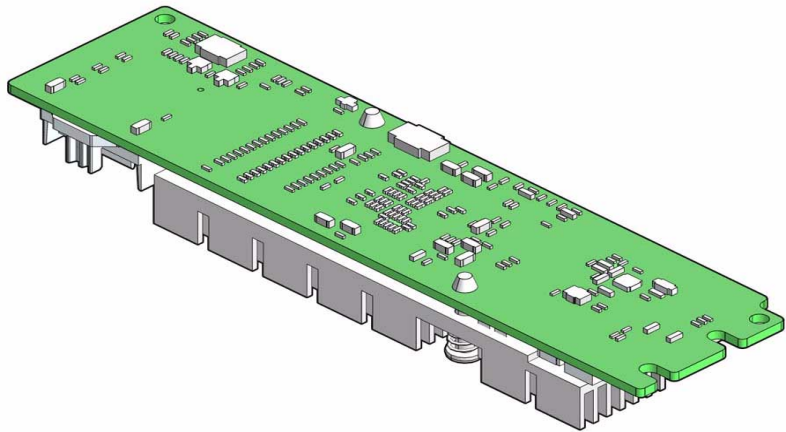
The Sun InfiniBand Dual Port 4x DDR PCIe Fabric Expansion Module (IB-FEM) is a single-wide InfiniBand DDR Fabric Expansion Module (FEM) based on Mellanox MT25408 ConnectX ASIC. The IB-FEM mezzanine card is installed on a compute blade to provide dual IB connectivity to the Sun Blade™ 6048 QDR Switched Network ExpressModule (IB-QNEM). The IB-FEM consists of one Amphenol NexLev connector and one Mellanox ConnectX ASIC chip. The NexLev connector contains a x8 PCIe and two DDR IB interfaces. This module is RoHS compliant. The IB-FEM is shown in [FIGURE 1-1](#).

---

**Note** – The Sun InfiniBand Dual Port 4x DDR PCIe Fabric Expansion Module is based on Mellanox Technologies’ MT25408 ConnectX IB adapter (HCA) device.

---

**FIGURE 1-1** Sun InfiniBand Dual Port 4x DDR PCIe Fabric Expansion Module



# Hardware and Software Requirements

Before installing the IB-FEM, ensure that your system meets the hardware and software requirements in [TABLE 1-1](#).

**TABLE 1-1** Hardware and Software Requirements for Sun x64 Servers

Requirements	Hardware or Software
Operating systems	<ul style="list-style-type: none"><li>• RedHat Enterprise Linux (RHEL) 5.3 and 4U7</li><li>• Suse Linux Enterprise Server (SLES) 10 update 2</li><li>• CentOS 5.3</li></ul>
Other software	<ul style="list-style-type: none"><li>• OpenFabrics Enterprise Distribution (OFED) 1.4 or later</li><li>• Mellanox OpenFabrics Enterprise Distribution (MLNX_OFED)</li></ul>
Network Express Module	Sun Blade 6048 InfiniBand QDR Switched Network Express Module (IB-QNEM)
Sun Blade Server Module	Sun Blade 6270 Server Module or other compatible server module



## Dual Port Connectivity

On some Sun Blade server modules, one port on the IB-FEM is disabled for compatibility with the IB-QNEM. Therefore two IB-FEMs are required for dual port connectivity. The Sun Blade X6270 Server Module requires that two IB-FEMs be installed on the server module for dual port connectivity to the IB-QNEM.

---

## Interfaces

### InfiniBand Interface

The IB-FEM is compliant with the *InfiniBand Architecture Specification, Release 1.2*. The IB-FEM has two compliant 4X InfiniBand ports, A and B. The IB-FEM provides access to these ports from the Sun Blade 6048 InfiniBand QDR Switched Network Express Module (IB-QNEM) through the chassis midplane. The IB-QNEM provides the InfiniBand connectors for external InfiniBand cables that are also compliant with the *InfiniBand Architecture Specification, Release 1.2*.

### PCI Express Interface

The Sun InfiniBand Dual Port 4x DDR PCIe Fabric Expansion Module has eight Tx/Rx pairs of SerDes, providing for a PCI Express x8 interface, version 2.0 compatible. The device can be either a master initiating the PCI Express bus operations or a slave responding to PCI bus operations.

---

## Memory

The IB-FEM support multiple memory devices through the PCI Express, Flash, and I<sup>2</sup>C compatible interfaces.

## System Memory

The IB-FEM uses the PCI Express interface to store and access IB fabric connection information on the system memory.

---

## Regulatory Compliance

Sun InfiniBand Dual Port 4x DDR PCIe Fabric Expansion Modules (also referred to as host channel adapters or HCAs) are components of the Sun server module in which they installed. See the Agency certifications of the applicable server module to determine the relevant Agency approvals for the server module and HCA combination.

To comply with legal EMI emissions regulations, the Sun Blade 6048 Modular Server must be provided with a front door assembly when used with the Sun InfiniBand Dual Port 4x DDR PCIe Fabric Expansion Module. The front door should remain closed during normal operation and opened only during product servicing or when power is removed from the product.

# Installing the InfiniBand Fabric Expansion Module

---

This chapter describes how to install the Sun InfiniBand Dual Port 4x DDR PCIe Fabric Expansion Module (IB-FEM) in your system and verify that it has been installed correctly.

This chapter includes the following section:

- [“Installing the IB-FEM” on page 5](#)

---

## Installing the IB-FEM

The following instructions describe the basic tasks required to install the IB-FEM inside a blade server module. Refer to your blade server’s system installation or server manual for detailed FEM instructions.

### ▼ To Install the IB-FEM

1. **Halt and power off your system.**
2. **Attach the adhesive copper strip of the antistatic wrist strap to the server module chassis. Wrap the other end twice around your wrist, with the adhesive side against your skin.**
3. **Perform an orderly shutdown of the server module.**
4. **Remove the server module from the system chassis.**
5. **Remove the cover from the server module.**

**6. Locate the FEM connectors on the server module.**

Ensure that you do use the correct connector. Look for the FEM label on the connectors.



---

**Caution** – Do not use the connectors labeled REM.

---

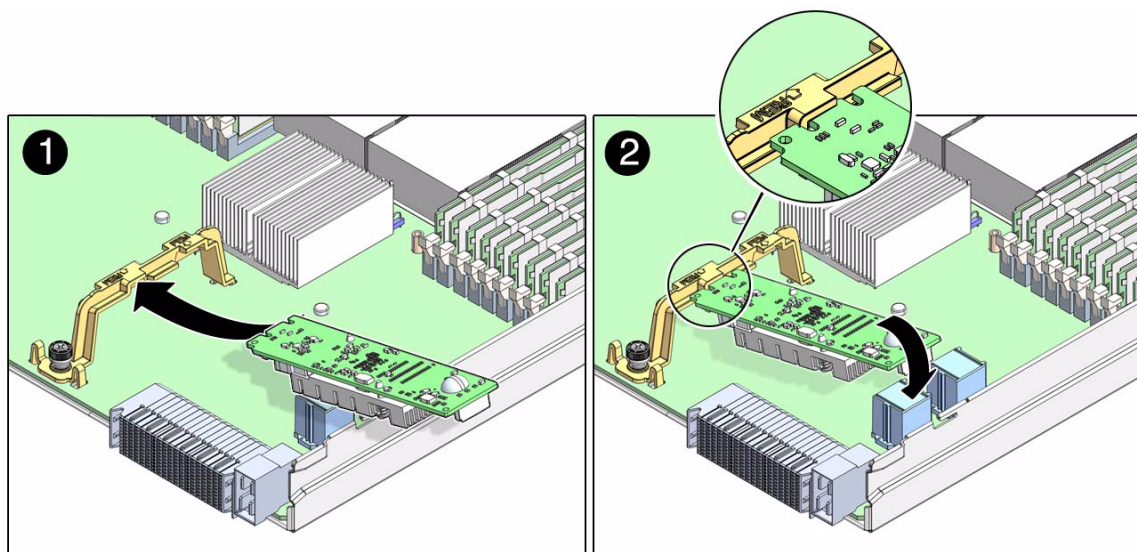
**7. Slide the IB-FEM card at an angle into the support bracket, then press it carefully into the connector.**

**8. Replace the cover on the server module.**

**9. Reinstall the server module into the chassis.**

Refer to the server module's service manual for specific instructions.

**FIGURE 2-1** Installing the IB-FEM



## InfiniBand Software on Linux

---

This chapter provides an overview and installation instructions for the InfiniBand software stack for the Linux operating system. Consult your Sun Blade server module product documentation for information about supported operating systems, firmware and software versions and updates, and other issues not covered in the main product documentation.

This chapter includes the following sections:

- [“InfiniBand Overview” on page 7](#)
- [“InfiniBand Software for Linux” on page 8](#)

---

## InfiniBand Overview

InfiniBand (IB) is a network architecture that is designed for the large-scale interconnection of computing and I/O nodes through a high-speed switched fabric. To operate InfiniBand on a Sun Blade 6048 Series Modular System (or other compatible Sun Blade Series Modular System), you need an Sun Blade X6048 QDR Switched Network ExpressModule (IB-QNEM) and an InfiniBand software stack.

# InfiniBand Software for Linux

If you have installed current releases of Red Hat Enterprise Linux Advanced Server (RHEL 5.3 or later) or SUSE Linux Enterprise Server (SLES10 update 2 or later) on a Sun Blade server module and you have installed the bundled drivers and OFED Release 1.4 or later, you do not need to install or configure additional drivers to support the IB-QNEM.

Specifically, RHEL contains support in the kernel for IB-HCA hardware produced by Mellanox (mthca driver). The kernel also includes core InfiniBand modules, which provide the interface between the lower-level hardware driver and the upper-layer InfiniBand protocol drivers. The InfiniBand modules provide user space access to InfiniBand hardware.

The kernel also includes the Sockets Direct Protocol (SDP) driver, IP over Infiniband (IPoIB) and the SCSI RDMA Protocol (SRP) driver.

RHEL includes the user space packages described in [TABLE 3-1](#).

**TABLE 3-1** RHEL Packages

kernel-ib	Base package that is required to support all other packages. Includes the files necessary to configure the kernel portion of the openib stack, create the proper udev rules, add the init script that allows the kernel modules to be selectively loaded at boot, and so on.
dapl	RDMA API that supports the DAT 1.2 specification.
ibibcm	InfiniBand Connection Management API.
libibcommon	Common utility functions for the IB diagnostic and management tools.
libibmad	Low-layer IB functions for use by the IB diagnostic and management programs, including MAD, SA, SMP, and other basic IB functions.
libibumad	User MAD library functions that sit on top of the user MAD modules in the kernel. Used by the IB diagnostic and management tools, including OpenSM.
libibverbs	Library that allows user space processes to use InfiniBand "verbs" as described in the InfiniBand Architecture Specification.
libibverbs-utils	Useful subnet and device diagnostic utilities.
libmthca	Device-specific user space driver for Mellanox HCAs (MT23108 InfiniHost and MT25208 InfiniHost III Ex) for use with the libibverbs library.
libipathverbs	Device-specific driver for Pathscale HCAs for use with libibverbs (only available on x86_64 and ia64 systems).
librdmacm	RDMA Connection Management (cm) library.

**TABLE 3-1** RHEL Packages

libsdp	Driver that enables a sockets application to use InfiniBand Sockets Direct Protocol (SDP) instead of TCP transparently and without recompiling the application.
openib-diags	Diagnostic programs and scripts that diagnose the IB subnet.
opensm	Subnet manager software for InfiniBand networks.
opensm-libs	Shared libraries for InfiniBand user space access.
perftest	InfiniBand performance tests.
srptools	In conjunction with the kernel <code>ib_srp</code> driver, enables discovery and use of SCSI class devices via the SCSI RDMA Protocol over InfiniBand.
flint	Tool to query and update firmware flash memory attached to Mellanox InfiniBand HCAs.

---

**Note** – These package names can change, depending upon the Linux OS.

---

The packages selected to support any given configuration will vary. [TABLE 3-2](#) lists the packages considered the absolute minimum needed to support the environment described in this guide.

**TABLE 3-2** Required Packages for InfiniBand Support

Package	Command Enabled	Description
kernel-ib	openibd	IB master control script
openib-diags	ibstat	IB utility to display HCAs
openib-diags	ibnetdiscover	IB utility to probe and show the fabric
flint	flint	Utility to update HCA FLASHRAM
libibcommon	NA	IB support package
libibmad	NA	IB support package
libibumad	NA	IB support package
OFED Release 1.4 or later	NA	IB support package

If you elected not to install these packages when installing the Linux OS or if you want to upgrade your drivers, you can install these packages at any time from the OS distribution source. You can also download the required files from [OpenFabrics.org](#). For information on both of these procedures, see [“Installing the InfiniBand Drivers on Linux” on page 11](#).

# OpenFabrics Enterprise Distribution for Linux

As the popularity of InfiniBand technology increases, the number of Linux distributions and open source organizations producing drivers and tools will increase. For up-to-date information, check with open source organizations and your current vendors.

The OpenFabrics organization is the Open Software solution in the InfiniBand software space. OpenFabrics Enterprise Distribution (OFED) is the InfiniBand suite of software produced by this organization. Various vendors contribute their drivers (and other software components) to OFED.

[TABLE 3-3](#) lists the tested Linux platforms and the corresponding OFED release.

**TABLE 3-3** Linux Platforms and OFED Release

Linux Platform	OFED Release
RHEL 4-U7	Sun has tested OFED Release 1.4 of the OpenFabrics stack.
RHEL 5.3	Note: You must install OFED Release 1.4 or a later version.
SLES10 SP2	Sun has tested OFED Release 1.4 for the SLES10 platform. Note: You must have OFED Release 1.4 or a later version.

OFED contains the following components:

- OpenFabrics core and Upper Layer Protocols (ULPs):
  - HCA drivers
  - Core
  - ULPs, including: IPoIB, SDP, SRP Initiator, iSER Host, RDS and uDAPL
- OpenFabrics utilities:
  - OpenSM (InfiniBand Subnet Manager)
  - Diagnostic tools
  - Performance tests
- Message Passing Interface (MPI):
  - OSU MPI stack supporting the InfiniBand interface
  - Open MPI stack supporting the InfiniBand interface
  - MPI benchmark tests (OSU BW/LAT, Pallas, Presta)
- Sources of all software modules (under conditions mentioned in the modules' LICENSE files)
- Documentation



# Installing the InfiniBand Drivers on Linux

If you did not install the InfiniBand drivers when installing the Linux OS, you can install them at any time from the OS distribution source or by downloading the necessary files from [OpenFabrics.org](http://OpenFabrics.org).

To do so, choose one of the following procedures:

- [“To Install IB Drivers From the Linux Distribution Source” on page 11](#)
- [“To Install the OFED Package” on page 13](#)

If you need to determine whether or not the drivers are already installed, see [“To Verify Driver Installation on Linux” on page 18](#).

## ▼ To Install IB Drivers From the Linux Distribution Source

### 1. Obtain the Red Hat Package Manager (RPM) files containing the InfiniBand drivers.

Access to these files is dependent on your individual installation configuration (net boot, CD/DVD boot, `.iso` files, and so on). When you decide on the appropriate access method and package selection, you can add the packages to the KickStart configuration file (on RHEL) for automatic inclusion in future installations.

---

**Note** – On a 32-bit RHEL 5.3 system, all packages have a `.i386.rpm` extension (as shown in the following procedure). On a 64-bit RHEL 5.3 system, all packages have a `.x86_64.rpm` extension instead.

---

### 2. Enter the `rpm -ivh` command for each InfiniBand package that you need to install.

Packages must be installed in the following order:

- `libibcommon`
- `libibumad`
- `libibmad`
- `openib-diags`
- `mstflint`

■ perfctest

The following example shows the installation of one package (libibcommon) and the resulting dialog on an RHEL 5.3 32-bit system:

```
> rpm -ivh libibcommon-1.0-1.i386.rpm
warning: libibcommon-1.0-1.i386.rpm: V3 DSA signature: NOKEY, key
ID db42a60e
Preparing...      ##### [100%]
1:libibcommon    ##### [100%]
> rpm -ivh libibumad-1.0-1.i386.rpm
.
.
.
```

3. If you are running the CSH or TCSH shell, enter the `rehash` command to rebuild the shell's view of available executables.

4. Enter the `ibstat` command to verify that the OS sees the IB-FEM.

```
> ibstat
CA 'mlx4_0'
    CA type: MT26418
    Number of ports: 2
    Firmware version: 2.6.0
    Hardware version: a0
    Node GUID: 0x0144f6c666bb1100
    System image GUID: 0x0144f6c666bb1103
    Port 1:
        State: Down
        Physical state: Polling
        Rate: 10
        Base lid: 0
        LMC: 0
        SM lid: 0
        Capability mask: 0x02510868
        Port GUID: 0x0144f6c666bb1101
    Port 2:
        State: Active
        Physical state: LinkUp
        Rate: 20
        Base lid: 8
        LMC: 0
        SM lid: 34
        Capability mask: 0x02510868
        Port GUID: 0x0144f6c666bb1102
CA 'mlx4_1'
    CA type: MT26418
    Number of ports: 2
```

```
Firmware version: 2.6.0
Hardware version: a0
Node GUID: 0x00144f6c6a300000
System image GUID: 0x00144f6c6a300003
Port 1:
    State: Down
    Physical state: Polling
    Rate: 2
    Base lid: 0
    LMC: 0
    SM lid: 0
    Capability mask: 0x02510868
    Port GUID: 0x00144f6c6a300001
Port 2:
    State: Active
    Physical state: LinkUp
    Rate: 20
    Base lid: 38
    LMC: 0
    SM lid: 34
    Capability mask: 0x02510868
    Port GUID: 0x00144f6c6a300002
>
```

5. (Optional) You can enter the `ibnetdiscover` command to verify the presence of an operational IB fabric.

For an example of the output of this command, see [“To Verify Driver Installation on Linux” on page 18](#).

6. (Optional) You can check the status of the `ib_ipoib` modules in the Linux kernel to determine whether the `ib_ipoib` driver is installed.

For details on this step, see [“To Install IPoIB Driver” on page 22](#).

## ▼ To Install the OFED Package

---

**Note** – The InfiniBand Dual Port 4x DDR PCIe Fabric Expansion Module requires OFED Release 1.4 or later.

---

1. On the Sun Blade server module, log in as `root` and copy the required files from the following location:

<http://www.openfabrics.org/downloads.htm>

In the following example, `OFED-1.4.tar` is used only as an example.

---

**Note** – You need write access to the files to execute the install script.

---

**2. From superuser (root), extract the files by typing:**

```
> tar -zxvf OFED-1.4.tar
```

**3. From the OFED-1.4 directory, initiate the installation process by typing:**

```
> ./install.sh
```

**4. When the InfiniBand OFED Distribution Software Installation menu appears, select option 2 (Install OFED Software).**

**5. When the Select OFED Software menu appears, select option 3 (All packages).**

**6. When you are asked if you want to create/install an MPI RPM with gcc, enter n.**

```
The following compiler(s) on your system can be used to
build/install MPI: gcc
Do you wish to create/install an MPI RPM with gcc? [Y/n]: n
```

**7. When you are asked if you want to create/install an openmpi RPM with gcc. Again, type n.**

```
The following compiler(s) on your system can be used to
build/install openmpi: gcc
Do you wish to create/install an openmpi RPM with gcc? [Y/n]: n
```

The installation script lists the OFED packages that it will build. See the following sample output.

```
Following is the list of OFED packages that you have chosen (some
may have been added by the installation program due to package
dependencies):
ib_ipath
ib_ipoib
...
mpitests
ibutils

WARNING: This installation program will remove any previously
installed IB packages on your machine.

Do you want to continue? [Y/n]: Y
```

**8. Type Y to continue.**

Next, you are prompted to configure InfiniBand IP support.

**9. Type Y when asked if you want to include IPoIB configuration files.**

```
Do you want to include IPoIB configuration files (ifcfg-ib*)?  
[Y/n]: Y
```

**10. Press Enter to accept the default when prompted to enter a temporary directory for OFED.**

```
RPM build process requires a temporary directory.  
Please enter the temporary directory [/var/tmp/OFED]:
```

**11. Press Enter to accept the default when prompted for the OFED installation directory.**

```
Please enter the OFED installation directory [/usr/local/ofed]:
```

At this point, the installer begins compiling InfiniBand packages. The process of building packages takes approximately 15–20 minutes.

The system displays output like the following:

```
The MPI_COMPILER_openmpi variable is not defined. Trying the  
default compiler: gcc  
  
The following compiler(s) will be used to build the openmpi RPM(s):  
gcc  
  
Checking dependencies. Please wait ...  
  
Building InfiniBand Software RPMs. Please wait...  
  
Building openib RPMs. Please wait...  
. . .  
33 packages were built  
  
Build process finished ...
```

Installation then begins. See the following message.

```
Removing previous InfiniBand Software installation
Running /bin/rpm -e libibverbs libibverbs-devel libibverbs-
utils...
```

The actual installation takes about one minute.

Assuming the IB-QNEM hardware is installed (and that an InfiniBand HCA is present), you are prompted to configure InfiniBand IP support.

**12. Enter Y in response to the following prompt:**

```
Do you want to configure IPoIB interfaces [Y/n]? Y
```

The default IPoIB interface configuration is based on DHCP. A special patch for DHCP is required for supporting IPoIB. The patch is available under:

OFED-1.4/docs/dhcp

If you do not have DHCP, you must change this configuration in the following steps.

The system displays the current configuration.

**13. When asked if you want to change the configuration as displayed, type y.**

```
The current IPOIB configuration for ibl is:
DEVICE=ibl
BOOTPROTO=dhcp
ONBOOT=yes
Do you want to change this configuration? [y/N]: y
```

The configuration script guides you through the changes one at a time. See the following as an example.

```
Enter an IP Address:10.0.0.52
Enter the Netmask: 255.255.255.0
Enter the Network:10.0.0.0
Enter the Broadcast Address:10.0.0.255
Start Device On Boot? [Y/n]:Y

Selected configuration:

IPADDR=10.0.0.52
NETMASK=255.255.255.0
NETWORK=10.0.0.0
BROADCAST=10.0.0.255
```

```
ONBOOT=yes
```

```
Do you want to save the selected configuration? [Y/n]: Y
```

**14. Type Y to save the configuration.**

If you have entered a valid IP configuration for `ib1`, you are now properly configured for IPoIB operations.

**15. Iterate the InfiniBand configuration over all InfiniBand interfaces.**

Enter a valid IP configuration for each network interface.

Once all IPoIB interfaces have been configured, you are prompted to configure OpenSM for the blade.

**16. Enter n to complete this part of the installation.**

```
Do you want to configure OpenSM [Y/n]? n
```

You should see a message like the following.

```
Installation finished successfully...  
Press Enter to continue...
```

**17. Press Enter.**

The InfiniBand OFED Distribution Software Installation Menu is displayed.

**18. Type Q to exit.**

The Sun Blade server module is configured now to start up the InfiniBand software on reboot (`ONBOOT=yes`).

If this is not the desired behavior, you can edit the `/etc/infiniband/openib.conf` file, specifying `ONBOOT=no`. You can also manually control basic InfiniBand behavior by entering the following command:

```
/etc/init.d/openibd option
```

where *option* can be `start`, `stop`, or `status`.

**19. After a successful installation, reboot the server module.**

After the reboot, the server module should boot as a functional member of the InfiniBand fabric.

## ▼ To Verify Driver Installation on Linux

1. Verify that the Linux software driver is installed and attached to the IB-QNEM by typing the `openibd status` command.

When using the `openibd` command, type the entire path as shown in this example.

```
> /etc/init.d/openibd status
HCA driver loaded

Configured IPoIB devices:
ib0 ib1 ib2 ib3

Currently active IPoIB devices:
ib1
ib3

The following OFED modules are loaded:

rdma_ucm
qlgc_vnic
ib_sdp
rdma_cm
ib_addr
ib_ipoib
ib_ipath
mlx4_core
mlx4_ib
ib_mthca
ib_uverbs
ib_umad
ib_sa
ib_cm
ib_mad
ib_core
iw_cxgb3
```

This example shows the IB driver installed, running and presenting one IB-HCA channel or network device (*ibn*) to the OS. In the example, the Linux network device appears as `ib1`. To view details of operational status, type the `ibstat` command.

The following example shows one operational IB channel into the IB fabric (or network). The `LinkUp` state indicates active participation in an IB fabric. The IB fabric is present as `lid 8` and it is being managed by `lid 34`.



```

> ibstat
CA 'mlx4_0'
    CA type: MT26418
    Number of ports: 2
    Firmware version: 2.6.0
    Hardware version: a0
    Node GUID: 0x0144f6c666bb1100
    System image GUID: 0x0144f6c666bb1103
    Port 1:
        State: Down
        Physical state: Polling
        Rate: 10
        Base lid: 0
        LMC: 0
        SM lid: 0
        Capability mask: 0x02510868
        Port GUID: 0x0144f6c666bb1101
    Port 2:
        State: Active
        Physical state: LinkUp
        Rate: 20
        Base lid: 8
        LMC: 0
        SM lid: 34
        Capability mask: 0x02510868
        Port GUID: 0x0144f6c666bb1102

```

You can also verify that the InfiniBand fabric is operational by entering the `ibnetdiscover` command. The output from this command will list all the nodes, as shown in the following sample output.

```

> ibnetdiscover
#
# Topology file: generated on Thu Jan 11 15:19:59 2007
#
# Max of 4 hops discovered
# Initiated from node 001b00000ca72620 port 001b00000ca72621

vendid=0x8f1
devid=0x5a31
sysimguid=0x8f10400411ef9
switchguid=0x8f10400411ef8

Switch 24 "S-0008f10400411ef8" # Switch port 0 lid 9
[21] "H-0002c90109761ea0" [2]
[12] "S-0005ad00000161ba" [5]

```

```

[7]          "H-001b00000ca72630" [1]
[6]          "H-001b00000ca72620" [1]
vendid=0x5ad
devid=0xa87c
sysimgguid=0x5ad01010161b6
switchguid=0x5ad00000161ba
Switch 8 "S-0005ad00000161ba"      # Switch - U3 port 0 lid 3
[4] "        H-0005ad0000011310" [1]
[3]          "S-0005ad00000161b6" [1]
[2]          "S-0005ad00000161b6" [2]
[1]          "S-0005ad00000161b8" [3]
[5]          "S-0008f10400411ef8" [12]
.
.
.
vendid=0x2c9
devid=0x6274
sysimgguid=0x1b00000ca72633
caguid=0x1b00000ca72630
Ca 1 "H-001b00000ca72630"      # 4x DDR IB 10-Port PCIe Network
Express Module
[1]          "S-0008f10400411ef8" [7]      # lid 68 lmc 0n

```

# Internet Protocol Over InfiniBand

---

This chapter describes configuration aspects of running the Internet Protocol over InfiniBand (IPoIB) on Linux.

This chapter includes the following sections:

- [“Configuring IPoIB on Linux” on page 21](#)
  - [“To Install IPoIB Driver” on page 22](#)
  - [“To Change IPoIB Configuration Without Rebooting” on page 23](#)
  - [“To Change IB Startup Behavior on Linux” on page 24](#)

---

## Configuring IPoIB on Linux

You might decide to change your IPoIB configuration for a variety of reasons, including the installation of an additional Sun Blade X6048 QDR Switched Network Express Module (IB-QNEM).

## ▼ To Install IPoIB Driver

1. **Determine whether the IPoIB driver is already installed by typing the `lsmod | grep ib` command.**

The output from this command shows all the IB drivers.

In the following sample output, note that the driver, `ib_ipoib`, is not listed.

```
> lsmod | grep ib
ib_sdp                45340 0
rdma_cm               26760 1 ib_sdp
ib_addr              10504 1 rdma_cm
ib_local_sa          14232 1 rdma_cm
findex               6528 1 ib_local_sa
ib_ipath             70552 0
ipath_core           179652 1 ib_ipath
ib_mthca             139184 0
ib_uverbs            47536 0
ib_umad              19888 0
ib_ucm               21512 0
ib_sa                18196 2 rdma_cm,ib_local_sa
ib_cm                39952 2 rdma_cm,ib_ucm
ib_mad               43176 5 ib_local_sa,ib_mthca,ib_umad,ib_sa,ib_cm
ib_core              59520 11
ib_sdp,rdma_cm,ib_local_sa,ib_ipath,ib_mthca,ib_uverbs,ib_umad,
ib_ucm,ib_sa,ib_cm,ib_mad
```

2. **To install the IPoIB driver, enter the `modprobe` command:**

```
> modprobe ib_ipoib
```

3. **Enter the `lsmod | grep ib` command again.**

Note that `ib_ipoib` is now listed.

```
> lsmod | grep ib
ib_ipoib             59800 0
ib_sdp               45340 0
rdma_cm              26760 1 ib_sdp
ib_addr              10504 1 rdma_cm
ib_local_sa          14232 1 rdma_cm
.
.
.
ib_core              59520 11
ib_sdp,rdma_cm,ib_local_sa,ib_ipath,ib_mthca,ib_uverbs,ib_umad,
ib_ucm,ib_sa,ib_cm,ib_mad
```

4. Type the `ifconfig` command to check for network interface `ib0`.

```
> ifconfig ib0
ib0      Link encap:UNSPEC  HWaddr 00-00-00-00-00-00-00-00-00-00-
00-00-00-00-00-00
          BROADCAST MULTICAST  MTU:2044  Metric:1
          RX packets:0 errors:0 dropped:0 overruns:0 frame:0
          TX packets:0 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:128
          RX bytes:0 (0.0 b)  TX bytes:0 (0.0 b)
```

Note that network interface `ib0` is present but has no valid IP address.

To assign an address, see [“To Change IPoIB Configuration Without Rebooting” on page 23](#).

## ▼ To Change IPoIB Configuration Without Rebooting



---

**Caution** – The changes made by this procedure are only temporary. These changes are lost during a reboot. To make permanent changes, see [“To Change IB Startup Behavior on Linux” on page 24](#).

---

1. Start the InfiniBand IP network by typing the `ifconfig` command and assigning a valid IP address for `ib0`.

```
> ifconfig ib0 10.0.0.50/24
ib0      Link encap:UNSPEC  HWaddr 00-00-00-00-00-00-00-00-00-00-
00-00-00-00-00-00
          inet
addr:10.0.0.50 Bcast: 10.0.0.255  Mask:255.255.255.0
          UP BROADCAST RUNNING MULTICAST  MTU:2044  Metric:1
          RX packets:0 errors:0 dropped:0 overruns:0 frame:0
          TX packets:0 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:128
          RX bytes:0 (0.0 b)  TX bytes:0 (0.0 b)
```

If `ib0` is unconfigured, it appears without an IP address as shown in the following output.

```
> ifconfig ib0
ib0      Link encap:UNSPEC  HWaddr 00-00-00-00-00-00-00-00-00-00-
00-00-00-00-00-00
          BROADCAST MULTICAST  MTU:2044  Metric:1
          RX packets:0 errors:0 dropped:0 overruns:0 frame:0
          TX packets:0 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:128
          RX bytes:0 (0.0 b)  TX bytes:0 ( 0.0 b)
```

## 2. Type the `route` command to verify that the 10.0.0 subnet is available.

The following output shows subnet 10.0.0 present and routed through `ib0`.

```
> route
Kernel IP routing table
Destination Gateway    Genmask         Flags   Metric Ref  Use  Ifac
e
10.0.0.0      *                   255.255.255.0   U        0      0    0   ib0
10.8.134.0    *                   255.255.255.0   U        0      0    0   eth0
169.254.0.0   *                   255.255.0.0     U        0      0    0   eth0
default       ban3rtr0d0 0.0.0.0         UG        0      0    0   eth0
```

## 3. As shown in the following example, you can enter the `ping` command to see another IPoIB node on the 10.0.0 subnet:

```
> ping 10.0.0.79
PING 10.0.0.79 (10.0.0.79) 56(84) bytes of data.
64 bytes from 10.0.0.79: icmp_seq=0 ttl=255 time=1.82 ms
64 bytes from 10.0.0.79: icmp_seq=1 ttl=255 time=0.082 ms
64 bytes from 10.0.0.79: icmp_seq=2 ttl=255 time= 0.062 ms

--- 10.0.0.79 ping statistics ---
3 packets transmitted, 3 received, 0% packet loss, time 2001ms
rtt min/avg/max/mdev = 0.062/0.655/1.823/0.825 ms, pipe 2
```

At this point, the IPoIB network is active and properly configured without rebooting.

# ▼ To Change IB Startup Behavior on Linux

You can change the InfiniBand startup behavior by editing the `openib.conf` configuration file. The full path is `/etc/infiniband/openib.conf`.

In the following example, `openib.conf` specifies that whenever the system boots, the InfiniBand services (IPoIB) and the SDP IP service are to start up automatically (`ONBOOT=yes`, `IPOIB_LOAD=yes`, `SDP_LOAD=yes`). However, `openib.conf` specifies that the SRP service is *not* to start up automatically (`SRP_LOAD=no`). You can alter any and all of these parameters.

### 1. Edit `ipoib.conf`.

The following is an example of the `ipoib.conf` file.

```
# Start HCA driver upon boot
ONBOOT=yes
# Load UCM module
UCM_LOAD=no
# Load RDMA_CM module
RDMA_CM_LOAD=no
# Load RDMA_UCM module
RDMA_UCM_LOAD=no
# Load MTHCA
MTHCA_LOAD=yes
# Load IPATH
IPATH_LOAD=yes
# Load IPoIB
IPOIB_LOAD=yes
# Load SDP module
SDP_LOAD=yes
# Load SRP module
SRP_LOAD=no
# Load RDS module
RDS_LOAD=no
```

## 2. Create (or edit) the `ifcfg-ibn` file to configure an individual network interface.

For each InfiniBand network interface, you need a corresponding startup file (`ifcfg-ibn`) in your startup scripts directory.

The directory for startup scripts might be `/etc/sysconfig/network/`, `/etc/sysconfig/network-scripts/`, or similar, depending on your version of Linux distribution.

As an example, the startup file for `ib0` might look something like the following.

```
more /etc/sysconfig/network-scripts/ifcfg-ib0
DEVICE=ib0
BOOTPROTO=static
IPADDR=10.0.0.50
NETMASK=255.255.255.0
NETWORK=10.0.0.0
BROADCAST=10.0.0.255
ONBOOT=yes
```

The `ONBOOT=yes` parameter indicates that the corresponding IP network interface is to automatically start up when the system boots. Specifying `ONBOOT=no` will configure the interface but not start it.

Generally, if you have enabled IPoIB services, the IB stack installation scripts automatically create the `ifcfg-ibn` configuration files for all IB network interfaces present. If you install an IB-QNEM *after* you have installed the IB stack, you might need to manually create the `ifcfg-ibn` files for the newly installed network interfaces.



# Specifications

This appendix includes the following section:

- [“IB-FEM Specifications” on page 27](#)

## IB-FEM Specifications

[TABLE A-1](#) lists the specifications for the IB-FEM card.

**TABLE A-1** Specifications for IB-FEM

<b>Physical</b>	
Size	5.5 in. x 1.278 in. (139.7 mm x 35 mm)
Airflow	200 LFM @ 70°Cmax.
<b>Protocol Support</b>	
InfiniBand	IBTA v1.2, autonegotiation* 5 Gb/s, 2.5 Gb/s
QoS	8 InfiniBand Virtual Lanes for each port
RDMA Support	Yes, all ports
<b>Power and Environmental</b>	
Voltage	12V, 3.3V
Maximum Power	12.5W
Operating Temperature	5°C to 35°C (for commercial servers) 0°C to 55°C (for NEBS-certified servers)

**TABLE A-1** Specifications for IB-FEM

<b>Regulatory</b>	See <a href="#">“Regulatory Compliance” on page 4.</a>
Safety	IEC/EN 60950-1:2001, ETSI EN 300 019-2-2
Environmental	IEC 60068-2- 64, 29, 32
RoHS	RoHS-R5

\* The autonegotiation protocol is proprietary of Mellanox Technologies and compliant with the *InfiniBand Architecture Specification, Release 1.2.*

# Index

---

## A

access to RPM files, 11

## C

commands

ibnetdiscover, 19

ifconfig, 23

lsmod, 22

modprobe, 22

openibd, 17

openibd status, 18

ping, 24

route, 24

configuring IPoIB, 16

controlling basic IB behavior, 17

## D

DHCP, patch required for default IPoIB, 16

## I

I/O

interfaces, 3

ibnetdiscover command, 19

ifcfg-ibn, 26

ifcfg-ibn file

configure IBolP, 15

ifconfig command, 23

ifconfig command (check for network interfaces), 23

InfiniBand

changing startup behavior, 24

configuration file (openib.conf), 17

controlling basic behavior, 17

installing the IB-FEM, 5

IPoIB configuration files, 15

## K

KickStart configuration file (RHEL), 11

## L

Linux platforms (tested with OFED 1.0), 10

lsmod command, 22

## M

modprobe command, 22

## N

network interface startup file (ifcfg-ibn), 26

## O

OFED

configuring IPoIB interfaces, 16

including IPoIB configuration files, 15

installation directory (default), 15

package components, 10

openib.conf file

full path, 24

openib.conf file (InfiniBand configuration file), 17

openibd command, 17, 18

## P

ping command (viewing the subnet), 24

## **R**

- Red Hat Package Manager files, access, 11
- RHEL kernel support for HCA hardware, 8
- `route` command (viewing subnet availability), 24
- RPM, access to Red Hat Package Manager files, 11

## **S**

- software stack, OFED, 10
- startup behavior (IB), changing, 24
- startup file, InfiniBand network interface, 26