

## **InfiniBand Hardware Installation and Cabling Guide Web Release**

Copyright © 2005 Topspin Communications, Inc. All rights reserved.

The Topspin Switched Computing System, Topspin Host Channel Adapter, Topspin Element Manager, and collateral software programs and documentation are subject to and made available only pursuant to the license agreement signed by you and Topspin, Communications, Inc. (or if no signed license agreement exists, the license agreement included with your original media) and may only be used in accordance with the terms of that agreement. Making copies, modifications, or compilation works of the software except as specifically allowed in that agreement is prohibited by law and constitutes a punishable violation of the law. This documentation may not be copied, modified or reduced to any electronic or machine-readable form without Topspin Communication, Inc.'s prior written consent.

As defined in FAR section 2.101, DFAR section 252.227-7014(a)(1) and DFAR section 252.227-7014(a)(5) or otherwise, the collateral software programs provided to you are “commercial items,” “commercial computer software” and/or “commercial computer software documentation.” Consistent with DFAR section 227.7202, FAR section 12.212 and other sections, any use, modification, reproduction, release, performance, display, disclosure or distribution thereof by or for the U.S. Government shall be governed solely by the terms of the license agreement and shall be prohibited except to the extent expressly permitted by the terms of that agreement. Any technical data provided that is not covered by the above provisions shall be deemed “technical data-commercial items” pursuant to DFAR section 227.7015(a) or equivalent. Any use, modification, reproduction, release, performance, display or disclosure of such technical data shall be governed by the terms of DFAR section 227.7015(b) or equivalent.

This documentation may include technical inaccuracies or typographical errors and is subject to correction and other revision without notice. TOPSPIN PROVIDES THIS DOCUMENTATION “AS IS” AND WITHOUT ANY WARRANTIES OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states or jurisdictions do not allow disclaimer of express or implied warranties in certain transactions; therefore, this statement may not apply to you.

©Copyright 2005, Topspin Communications, Inc. All rights reserved. Topspin is a registered trademark and the Topspin logo, TopspinOS, and Topspin Switched Computing System are trademarks of Topspin Communications, Inc. Other company, product, or service names are the property of their respective owners.

Document Version: Web

Part Number: 10-00122-WEB

April, 2005

Printed in the United States of America.

# Contents

## Preface ..... vii

Intended Audience .....	vii
Typographical Conventions .....	vii
Contact Information .....	viii

## 1: Connecting the Hardware ..... 1

Introduction .....	1
Definitions .....	3
Cluster Component Naming Conventions .....	3
Node Naming .....	4
Rack Naming .....	4
Node Network Interfaces .....	4
..... Primary InfiniBand ib0 Interface - Naming and Address Assignment	5
Ethernet Management eth0 Interface - Naming and Address Assignment .....	5
Backup InfiniBand ib1 Interface - Naming and Address Assignment .....	5
Backup Ethernet eth1 Interface - Naming and Address Assignment .....	6
Leaf Switch Naming and Address Assignment .....	6
Core Switch Naming and Address Assignment .....	6
TS-360 Edge switches and Address Assignment .....	6
Cable Mapping and Labeling .....	6
General Cable Considerations .....	7
Server Nodes to Leaf Switch Ports .....	7
Leaf Switch to Core Switch ports .....	7
Installation Task and Timing Overview .....	9
Installation Activities .....	9
InfiniBand cable management .....	10
Installing and Racking Switches .....	13

Initialize all Leaf and Core switches .....	14
IB cable lay-down from Leafs to Cores .....	15
Connect Leaf switches to Core .....	16
Connect Nodes to Leaf switches .....	16
Install Host-side Software .....	16
Power on Cluster .....	17
Troubleshoot Bad Cables .....	17
Installation Risks and Issues .....	17
IB Core Cable Mass .....	17
Spares .....	20

## 2: Getting Started ..... 21

Introduction .....	21
Overview of Set-up Procedures .....	22
Install HCAs and Drivers in Hosts .....	22
Install and Power on the InfiniBand Chassis .....	22
Configure Basic Connectivity .....	22
Test Network Connectivity .....	22
Verify Communication Between Hosts .....	23
Verify Switch to Host Communication .....	23
Install the GUI (Element Manager) .....	23
Access a Management System .....	23
CLI Management .....	23
GUI Management .....	24
SNMP Management .....	24
Configure Partitions .....	24
Create a Partition ID (P_Key) .....	24
Specify Partition Members and the Membership Type .....	26
Maintain Partition Key Information .....	26
Set User Levels and Passwords .....	27
Change Default User Name and Password .....	27
(Optional) Configure any Ethernet Gateways .....	28
(Optional) Configure any Fibre Channel Gateways .....	28
(Optional) Configure the VFrame Director .....	28

## 3: Cable Length Sizing ..... 29

Sizing .....	29
--------------	----

## 4: Physical and Environmental Specifications 31

Specifications .....	31
----------------------	----

## 5: P-Touch Editor Notes..... 33

Introduction.....33

## 6: IB Port Test ..... 35

Introduction.....35

Script.....35



# Preface

---

This document is a guide to the TopspinOS Command Line Interface, or CLI. This document explains how to use the TopspinOS CLI and provides a categorized, alphabetical listing all available CLI commands.

## Intended Audience

Topspin provides this document for administrators who install, configure, and manage Topspin equipment. This document assumes that administrators have prior Ethernet, Fibre Channel, and network administration experience.

## Typographical Conventions

The following typographic conventions are used in this manual to provide visual clues as to the purpose or application of specific text.

- **Bold** text indicates a command or keyword, or text that appears in your display.
- *Italics* indicate variables that you replace with an actual value.
- Square brackets ([,]) indicate an optional argument that you choose to include or exclude when you enter a command.
- Pipe character (|) indicates an “or” choice. For example, “**a | b**” indicates “a or b.”
- Ellipses (...) indicate truncated text. You will see these in long examples depicting terminal output that is too long to be shown in its entirety.



---

**NOTE:** Indicates an important point or aspect that you need to consider before continuing.

---

# Contact Information

**Table III-1:** Customer Contact Information

Address	Topspin Communications, Inc. 515 Ellis St. Mountain View, CA 94043
Telephone, Corporate Headquarters	(866) TOPSPIN or (650) 316-3300
Fax	(650) 316-3269
Telephone, Technical Support	(800) 499-1473
Email, Technical Support	support@topspin.com
Web site, Support	<a href="http://support.topspin.com">http://support.topspin.com</a>



# Connecting the Hardware

---

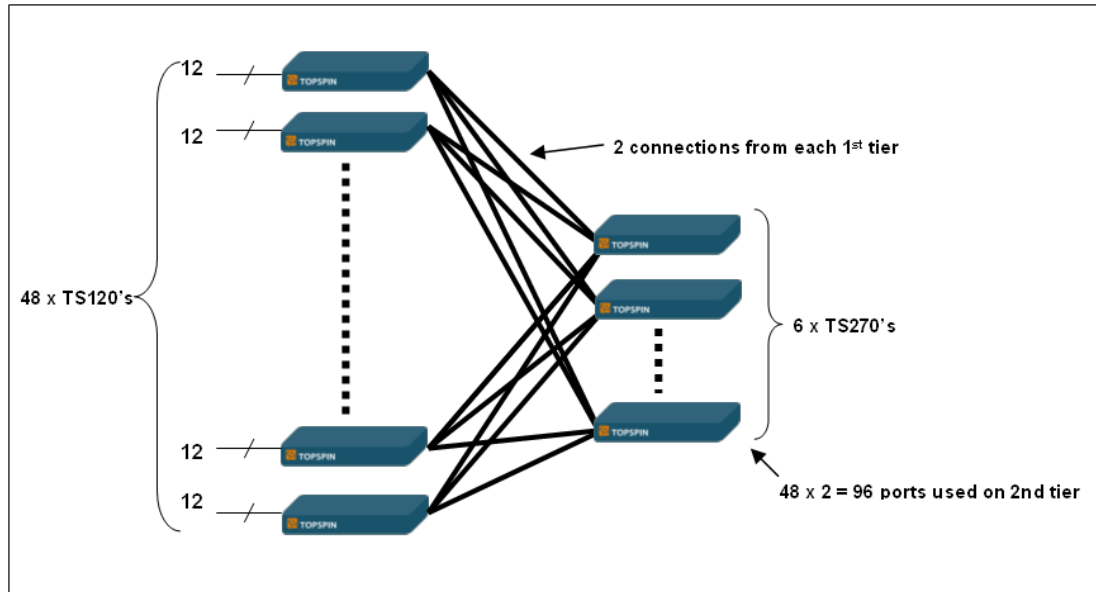
The following sections appear in this chapter:

- [“Introduction” on page 1](#)
- [“Definitions” on page 3](#)
- [“Cluster Component Naming Conventions” on page 3](#)
- [“Cable Mapping and Labeling” on page 6](#)
- [“Installation Task and Timing Overview” on page 9](#)
- [“Installation Activities” on page 9](#)
- [“Installation Risks and Issues” on page 17](#)

## Introduction

The purpose of the document is to establish the Best Practices and Guidelines in developing an architecture, sizing, installing and managing large InfiniBand clusters using Topspin Server Switches.

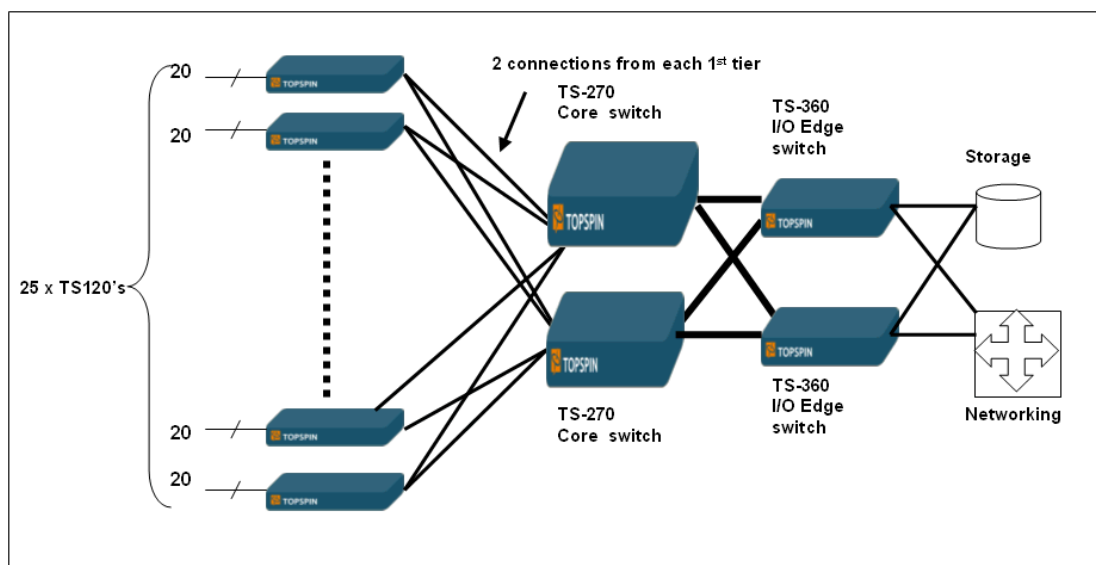
A sample InfiniBand Fabric topology diagram for a High Performance Computing (HPC) cluster might look like the following:



**Figure 1-1:** Example HPC Cluster

The diagram in [Figure 1-1](#) illustrates the topology for a 576-node non-blocking High Performance Computing (HPC) cluster. Non-blocking topologies provide, for each path into a switch or network, an equal path out. Every 12 nodes would connect directly to a TS-120 leaf switch. Each TS-120 switch would connect with two ports to each of 6 TS-270 core switches.

A sample InfiniBand Fabric topology diagram for a Data Center (DC) cluster might look like the following:



**Figure 1-2:** Example Data Center Cluster

The diagram in [Figure 1-2](#) illustrates the topology for a 500-node blocking Data Center (DC) cluster. Every 20 nodes would connect directly to a TS-120 leaf switch. Each TS-120 switch would connect with two ports to each of 2 TS-270 core switches. The TS-270's would then be cross connected to 2 TS-

360 I/O Edge Switches. The TS-360 Edge Switches have Ethernet and Fibre Channel IB Gateways which cross connect to external storage and networking subsystems.

## Definitions

**Table 1-1:** InfiniBand Terms and Definitions

Term	Definition
Cluster Node Number	A number that uniquely identifies every node in a cluster. For example, a 500-node cluster would have Cluster Node Numbers from 1 to 500.
CLI	The Command Line Interface (CLI) for the switch.
Core Switch	Core Switches form the second tier of an InfiniBand fabric and are used to interconnect Leaf Switches and Edge Switches. They form the backbone of the InfiniBand fabric. A typical Core Switch would be a 96-port Topspin TS-270.
Edge Switch	Edge Switches provide support for InfiniBand I/O Gateways and are used to connect an InfiniBand fabric to external network and storage subsystems. A typical Edge Switch would be a 24-port Topspin TS-360 with support for up to 12 Gigabit Ethernet or Fibre Channel Gateways.
Host Channel Adapter (HCA)	An InfiniBand Host Channel Adapter (HCA) connects the Host to the InfiniBand Fabric. HCAs are either PCI-X or PCI-Express based. They are typically dual ported to allow for redundant InfiniBand connections. Each port runs at 10 Gb/s.
HPC	High Performance Computing
InfiniBand Fabric	A collection of interconnected InfiniBand Nodes, Leaf switches and Core switches that form a continuous data network. Also referred to as a “Cluster”.
Leaf Switch	Leaf Switches form the first tier of an InfiniBand Fabric and are used to connect Hosts to the Fabric. A typical Leaf Switch would be a 24-port Topspin TS-120.
Management Network	Also referred to as “Administration Network”. A 100 Base T or Gigabit Ethernet network used for out-of-band administration and management of the nodes and switches.
Node	A single compute element of the cluster, for example a 1U Server with one or more CPUs
Rack Node Number	A number that uniquely identifies a node within a rack or frame. Typically a Rack Node Number will range from 1 (lowest slot in a rack) to 42 (for a 42 RU Rack).

## Cluster Component Naming Conventions

It is extremely important that a consistent device naming and network address numbering scheme be adopted before the installation of any large InfiniBand fabric. This section describes the recommended naming conventions for all components in the cluster.

## Node Naming

All nodes should be named in a way that uniquely identifies the node within the cluster. There are a variety of ways to do this. Namely:

<b>Numerical Naming</b>	A Name based on Cluster Size, i.e. Node123
<b>Physical Naming</b>	A Name based on Physical Rack Location (e.g. Node R3S22 = Rack 3 - Slot 22)
<b>Topological Naming</b>	A name based on Cluster Topology Location (e.g. Node L7P10 = Leaf Switch #7 - Port #10)

There are advantages and disadvantages to each. Certainly, **Numerical Naming** is the simplest, but provides no other information about the node. **Physical Naming** aids in any installation, cabling or maintenance activities. **Topological Naming** aids in assigning nodes in an optimal fashion for very high performance applications, as nodes that share leaf switches will in general experience higher performance communication than nodes that are separated by an entire cluster fabric.

Regardless of what Node Name philosophy is adopted, all of the information about a node's physical and topological information will need to be collected in a Cluster Information Table for cross reference purposes.

The following sections describe each of these naming methods in detail:

Numerical Naming	The format of the Node Name should be a fixed prefix followed by the node's Cluster Node Number, a globally unique number starting with 1 and increasing numerically. Nodes within a given rack should be numbered so that the node numbering increases from bottom slots to top slots, corresponding with rack labeling for a standard 42U rack. <b>Example:</b> The 42nd Node = Node42
Physical Naming	The format of the Node Name should be a fixed prefix followed by the node's Rack Number, followed by the Rack Slot Number (1-44 increasing from bottom to top). <b>Example:</b> Node located at Rack 3, Slot 22 = NodeR3S22
Topological Naming	The format of the Node Name should be a fixed prefix followed by the node's Leaf Switch Number, followed by the Switch Port Number. <b>Example:</b> Node located at Leaf Switch #7, Slot #10 = NodeL7S10

## Rack Naming

Racks should be named numerically in increasing order regardless of the number of rows of racks. This is because the rack number will be used to calculate the IP address of nodes located within the rack, and therefore rack numbers should range from 0 to 254. The term “frame” and “rack” can be used interchangeably. The format of the Rack Name should be a fixed prefix followed by the rack's number, a unique number starting with 1 and increasing numerically, for example “rack7”.

## Node Network Interfaces

Each node has access to two types of networks, the InfiniBand fabric used for application and data transfers, and the out-of-band Ethernet management network. In addition, each node may be set up with a primary connection and an optional back-up connection for either the InfiniBand or the Ethernet interfaces. This means that nodes can have up to four external network interfaces: They are named as follows:

- ib0: Primary InfiniBand network
- ib1: Back-up InfiniBand network

- eth0: Primary Ethernet management network
- eth1: Back-up Ethernet management network

Each interface should have a unique hostname and IP address according to the following convention.

## Primary InfiniBand ib0 Interface - Naming and Address Assignment

The ib0 interface will be used as the primary high-performance network interface. The hostname for the InfiniBand interface will be the same as the node's Node Name.

### Example

The InfiniBand hostname for “**node42**” would be “**node42**”.

The IP address for the InfiniBand interface should be assigned in the **10.0.X.X** subnet with a subnet mask of **255.255.0.0** and having an IP address determined as follows:

**10.0.rack\_Number.rack\_Node\_Number**

Where **rack\_Slot\_Number** corresponds to the slot number or node number of the node within the rack.

### Example

If we assume **node42** is in Rack Number 2, 6th node from the bottom, then the IP address for the ib0 interface on “**node42**” is **10.0.2.6**.

## Ethernet Management eth0 Interface - Naming and Address Assignment

The eth0 interface will typically be reserved for administration and management of the node. The hostname for the eth0 interface will be of the form “**NodeName-eth0**”.

The IP Address for the eth0 interface should be assigned in the **10.1.X.X** subnet with a subnet mask of **255.255.0.0**. The IP address should be of the form:

**10.1.rack\_Number.rack\_Node\_Number**

### Example

If we assume **node42** is in rack number 2, 6th node from the bottom, then the hostname and IP address for the eth0 interface would be “**node42-eth0**” and **10.1.2.6**.

## Backup InfiniBand ib1 Interface - Naming and Address Assignment

The **ib1** interface will be used as the backup high-performance network interface. The hostname for the ib1 InfiniBand interface will be of the form “**NodeName-ib1**”.

The IP address for the ib1 interface should be assigned in the **10.2.X.X** subnet with a subnet mask of **255.255.0.0** and have the same two lower-order octets as the ib0 interface.

### Example

If we assume **node42** is in rack number 2, 6th node from the bottom, then the hostname and the IP address for the ib1 interface would be “**node42-ib1**” and **10.2.2.6**.

## Backup Ethernet eth1 Interface - Naming and Address Assignment

The **eth1** interface will typically be reserved for failover scenarios when there has been a failure on the primary management network. The hostname for the **eth1** interface will be of the form “**NodeName-et1**”.

The IP Address for the **eth1** interface should be assigned in the **10.3.X.X** subnet with a subnet mask of **255.255.0.0** and have the same two lower-order octets as the **ib0** interface.

### Example

If we assume **node42** is in rack number 2, 6th node from the bottom, then the hostname and the IP address for the **eth1** interface would be “**node42-eth1**” and **10.3.2.6**.

## Leaf Switch Naming and Address Assignment

Each Leaf switch should have a hostname and IP address for access through the management interface. The hostname should be of the form “**leafX**”, where X is the number of the leaf switch, starting with 1. The IP address for the management interface for all Leaf switches should be of the form **10.4.0.leafX** with a subnet mask of **255.255.0.0**.

### Example

TS-120 Leaf Switch, “**leaf23**”, will have an IP address of **10.4.0.23**

## Core Switch Naming and Address Assignment

Each Core Switch should have a hostname and IP address for access through the management interface. The hostname should be of the form “**coreX**”, where X is the number of the core switch, starting with 1. The IP address for the management interface for all Core switches should be of the form **10.4.100.coreX** with a subnet mask of **255.255.0.0**.

### Example

TS-270 Core Switch, “**core4**”, will have an IP address of **10.4.100.4**.

## TS-360 Edge switches and Address Assignment

Each TS-360 Edge switch should have a hostname and IP address for access through the management interface. The hostname should be of the form “**edgeX**”, where X is the number of the edge switch, starting with 1. The IP address for the management interface for all Edge switches should be of the form **10.4.200.edgeX** with a subnet mask of **255.255.0.0**.

### Example

TS-360 Edge Switch, “**edge2**”, will have an IP address of **10.4.200.2**.

## Cable Mapping and Labeling

This section describes the mapping and cable routing scheme for:

- Server nodes to 1st tier Leaf switches
- 1st tier Leaf switches to 2nd tier Core switches
- 2nd tier Core switches to 3rd Edge switches

## General Cable Considerations

- Correctly sizing cables for a particular installation is extremely important. Please refer to “[Cable Length Sizing](#)” on page 29 for help in determining proper cable lengths.
- In general cables from all switches should be distributed evenly through both the right and left sides of a Rack.
- The TS-270 Server Switch has Line Interface Modules (LIMs) that number the ports from left to right, when looking at the ports on the rear of the chassis.
- The TS-120 has two rows of ports: ports 1-12 (bottom row, left to right) and ports 13-24 (top row, left to right).




---

**NOTE:** All cables destined for ports 1-6 or 13-18 of a TS-120 switch should be routed through the left side of the rack. All cables destined for ports 7-12 or 19-24 of a TS-120 switch should be routed through the right side of the rack, if possible.

---

- The TS-270 switch has 96 ports, organized as 8 slots of 12-port LIMs. Each LIM in the TS-270 is numbered from 1-8, top to bottom. The ports within each LIM are all numbered 1-12, left to right.




---

**NOTE:** All cables destined for ports 1-6 for all LIMs of a TS-270 switch should be routed through the left side of the rack. All cables destined for ports 7-12 of a TS-270 should be routed through the right side of the rack, if possible.

---

## Server Nodes to Leaf Switch Ports

TS-120 switches can be configured to have 24 4x ports, grouped in 2 rows of 12. (Alternative hardware configurations depend on port speeds). Port numbering is 1-12 for the top row of ports and 13-24 for the bottom row of ports. For non-blocking configurations, all Host nodes will connect to the TS-120 through the bottom set of 12 ports (port numbers 13-24). Each group of 12 nodes will connect to ports 13-24 in corresponding order.

### Example

```
node 240 -> port 13
node 241 -> port 14
...
node 251 -> port 24
```

All TS-120 Leaf switches will connect to the TS-270 Core switches through the top set of the TS-120's 12 ports (port numbers 1-12).

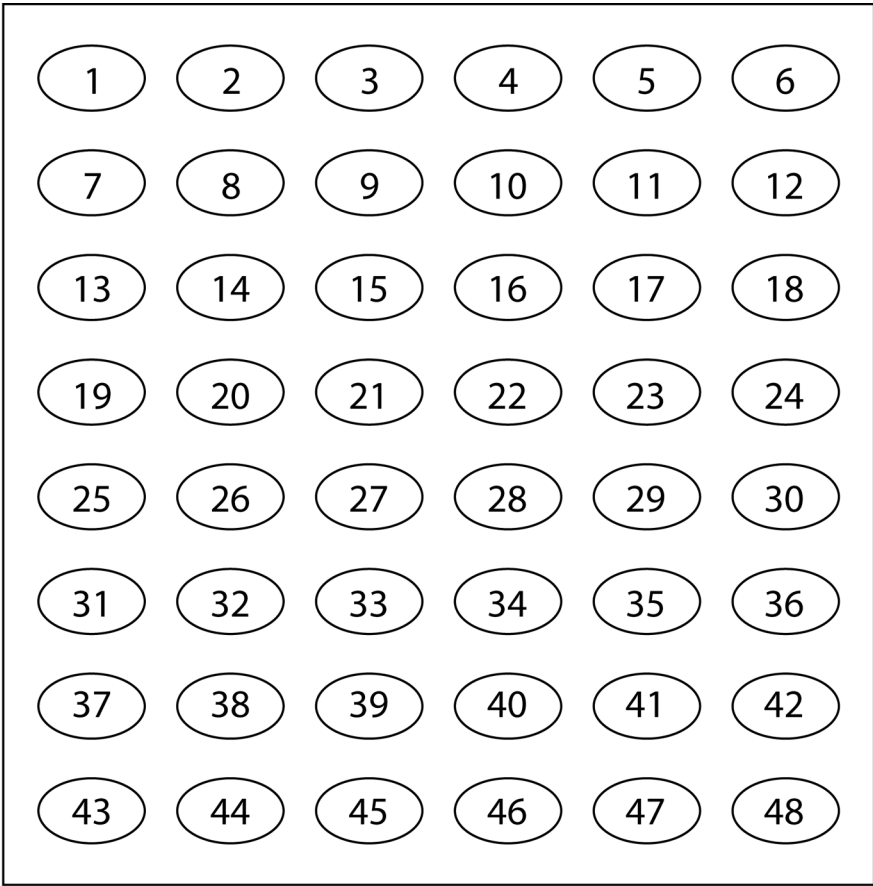
If labeling is done for node-leaf cables, then the following labeling scheme is recommended:

```
node# leaf#/port#
```

## Leaf Switch to Core Switch ports

In a distributed core topology, which should be typical of most HPC or data-center clusters, each TS-120 Leaf Switch will have the same number of ports connecting to each of the TS-270 Core Switches. In a non-blocking HPC cluster there would be 12 host and 12 core connections per Leaf Switch. In a typical DC cluster, there might be 20 host and 4 core connections per Leaf Switch (5:1 blocking).

Using the topology diagram from [Figure 1-1](#) as a model, each of the 48 TS-120 Leaf Switches would map to the same 2 ports for each of the TS-270 6 Core Switches as per the table below.

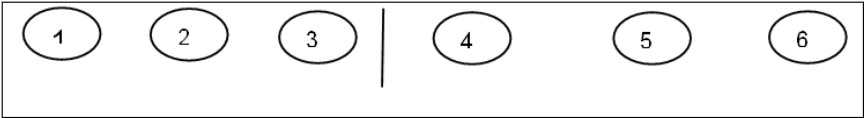


LIM port numbers

1,2      3,4      5,6      7,8      9,10      11,12

**Figure 1-3:** TS-270 Core Ports

Each of the Leaf switch ports would always map to the same Core switches as per the table below:



**Port numbers**    1 2    3 4    5 6    7 8    9 10    11 12

**Figure 1-4:** TS-120 Leaf Ports

Labeling all cables between the leaf and the core switches is highly recommended. Failure to label the leaf-core cables will hamper best efforts to isolate, identify and troubleshoot any bad cables when the cluster is brought up. Failure rates for InfiniBand cables in HPC installations have been measured at about 1% of all cables used.



# Installation Task and Timing Overview

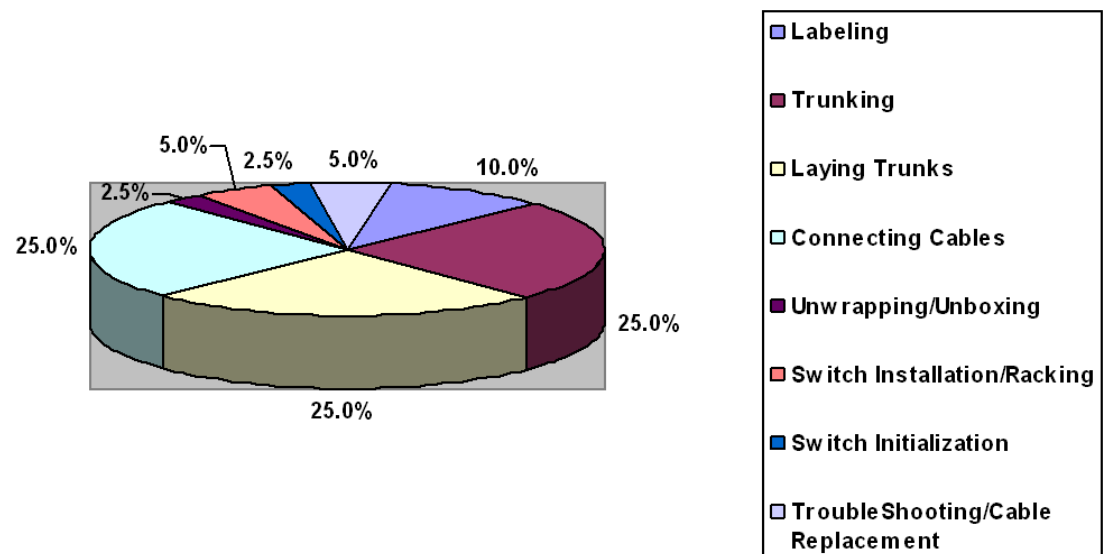
The amount of time and man-power required for installation will vary directly with the size of the cluster. As an example, a 468-node cluster took approximately 8-10 man-weeks to bring up the InfiniBand fabric. However, this example was an unusually challenging scenario because of the following factors:

- All racks were densely populated, with most utilizing 41 of 42 U of space.
- Installation of the leaf switches were done after the racks had been populated with nodes and internal management cabling, leaving very little free working space.
- Leaf switches were forced to be installed in a manner that greatly limited accessibility, both for switch racking within the rack and for connecting cables.

As a general guideline 85% of the installation time will be spent performing tasks associated with cable management, including:

- Cable labeling
- Creating cable trunks from the leaf to core switches
- Laying cable trunks between the leaf and core switches
- Connecting all InfiniBand cables to nodes and switches

The following chart illustrates the approximate task breakdown as a percentage of the overall installation.



**Figure 1-5:** Large Cluster Installation - Time Breakdown



**NOTE:** Unexpected issues are certain to arise, and installation complexity is certain to grow with the size of a cluster, regardless of previous experience or expectations.

## Installation Activities

This section focuses on the following topics:

- [“InfiniBand cable management” on page 10](#)
- [“Installing and Racking Switches” on page 13](#)

- [“Initialize all Leaf and Core switches” on page 14](#)
- [“IB cable lay-down from Leafs to Cores” on page 15](#)
- [“Connect Leaf switches to Core” on page 16](#)
- [“Connect Nodes to Leaf switches” on page 16](#)
- [“Install Host-side Software” on page 16](#)
- [“Power on Cluster” on page 17](#)
- [“Troubleshoot Bad Cables” on page 17](#)

## InfiniBand cable management

The most time-consuming aspect of the installation will be in cable management, including labeling and trunking leaf-core cables.

### Inventory all cable counts

Make sure there are sufficient cables of appropriate lengths to complete the installation.

### Prototype worst case path

Make sure that cables are of sufficient length to cover the worst case path in terms of cable length required. Validate previous assumptions by actually running a sample cable from the worst case distances between leaf and core switches, with respect to cable length anticipated. Make any plans for final provisioning of cables based on cable lengths actually needed.

### Unwrapping cables

Cables should be unrolled and laid flat down a long runway of floor space. Cables cannot be unrolled in place. They need to be walked into an unrolled flat position. Use a “hand over hand” method for unrolling the cable coil as you walk across the floor, such that the cable lies as flat as possible. Unwrap and label manageable portions of cables at a time.

### Cable Labeling System

- The recommended labeling and coordinate system for each type of cables is shown below in [Table 1-2](#). Under this scheme, each cable can be easily identified with respect to both its source and destination.

**Table 1-2:** Cable Labeling Scheme

Cable Type	End-1	End-2
Node - Leaf	1. NodeName 2. leaf# / LIM# /port#	1. leaf# / LIM# /port# 2. NodeName
Leaf - Core	1. leaf# / LIM# / port# 2. core# / LIM# / port#	1. core# / LIM# / port# 2. leaf# / LIM# / port#
Core - Edge	1. core# / LIM# / port# 2. edge# / LIM# / port#9	1. edge# / LIM# / port#9 2. core# / LIM# / port#

- Use 3/4” labels and have labels lay parallel with respect to the cable. If possible, have labels printed in a mirror fashion, such that the label can be read from either facing left or facing right perspective.
- Use **Brother P-touch Editor version 3.2**. This allows importing of an Excel CSV format database to drive the labeling process. Print labels in groups of 12 to keep manageable. Be sure to use tape that allows for easy peel across the back median of the label. Please refer to [“Cable Length Sizing” on page 29](#).

- Label the cables while still coiled. Do not unroll the cable first. Make sure both ends have the exact same label.
- Label all cables from the perspective of someone facing the back of a switch. They should all be oriented the same way, facing the reader who is facing the switch ports. Place the label about 6" from the base of the connector. Don't label right up close to the connector.

## **Trunking or Bundling**

Bundle IB cables in groups equal to the number of nodes and/or the number of leaf-core connections per leaf switch. For non-blocking configurations, this number will generally be 12.

Cover both ends of the cable trunks with bubble-wrap or a plastic bag to guard the connectors until they're connected, as illustrated in the following picture:



**Figure 1-6:** Example of Cable Trunking Setup

The core facing set of cables within a cable trunk should be further grouped and tie-wrapped according to the number of core racks. For example, if all core switches are racked into 2 racks, then each leaf-core cable trunk should be tie-wrapped into 2 groups such that all cables destined for the same core rack are grouped together. This only needs to be done for the core-facing set of cables, not the leaf-facing cables. This makes for an easier process of partitioning cables that are destined for the core switches, regardless of whether leaf-core cabling is laid under a raised floor or laid in overhead rails.

## Supplies to Have On Hand

In general it is highly recommended that the following supplies be on hand for any installation:

- Sufficient supplies of double-sided ½" Velcro, either in bulk or precut, for tie-wrapping cables together.
- Sufficient supplies of TZ easy-peel labeling tape
- Screwdrivers, both Phillips and Standard
- Laptop computer

## Installing and Racking Switches

- In general it is very highly recommended that all leaf and core switches be installed within the racks before shipment, such that unpacking and racking of switches does not have to be done as part of the on-site installation activity.
- It is also very highly recommended to have no more than two TS-270 core switches mounted in a single rack, due to the difficulty of maintaining working service space.
- It is also very highly recommended that all leaf switches be mounted, such that all cabling can be easily managed. For densely populated racks, it may be desirable to mount leaf switches flush-rear to the rack, so that all switch ports are easily accessible. For racks that are not densely populated, the leaf switches could be mounted flush-front to the rack, so long as all ports are easily accessible.



**NOTE:** TS-270 core switches may have certain line modules (LIMs) blanked out with blanking panels, for reduced port-count configurations. If so, ensure that the lowest numbered LIMs are populated and the highest numbered LIMs are blanked out, in order to be consistent with the labeling scheme (core# / LIM# / port#). Otherwise, disregard this note.

---

## Initialize all Leaf and Core switches

This activity can be done independent of all previous activities for all Leaf Switches. Unpack and stack all Leaf Switches one on top of another (in the case of TS-120 Leaf Switches), which makes for a much more streamlined switch initialization process, as seen in the picture below:



**Figure 1-7:** Leaf Switch Initialization

Initialize switches as follows (see the CLI reference manual for full syntax):

```

Login: super
Password: xxxxxx

Topspin-360> enable
Topspin-360# configure terminal
Topspin-360 (config)# hostname rack1-360-2
Topspin-360 (config)# interface mgmt-ethernet
rack1-360-2 (config-if-mgmt-ethernet)# ip address 10.3.0.2 255.255.0.0
rack1-360-2 (config-if-mgmt-ethernet)# gateway 10.3.0.1
rack1-360-2 (config-if-mgmt-ethernet)# no shutdown
rack1-360-2 (config-if-mgmt-ethernet)# exit
rack1-360-2 (config)# no ib sm subnet-prefix fe:80:00:00:00:00:00:00
rack1-360-2 (config)# exit
rack1-360-2# copy running-config startup-config
rack1-360-2# reload

```

This above sequence should be used for initializing all switches except the primary and secondary core. In these cases, omit step 3 (disabling the subnet manager). Instead, the primary core should run the subnet manager with the highest priority as follows:

**ib sm subnet-prefix fe:80:00:00:00:00:00:00 priority 30**

The secondary core should run the subnet manager, but at a lower priority:

**ib sm subnet-prefix fe:80:00:00:00:00:00:00 priority 10**

In general, the boot sequence for switches should begin with the primary core switch, since a previously running subnet manager with a lower priority will not preempt an existing running subnet manager with a higher priority.

## IB cable lay-down from Leafs to Cores

Cable trunks may need to be split/combed above the core racks, using overhead rail tracks as the comb, according to the leaf-core rack groups described in section [“Trunking or Bundling” on page 11](#). This is particularly significant where multiple racks are used for housing core switches.

Be sure to take up cable slack (or looseness) at any opportunity. Typically, the area around the core racks tends to be highly congested with cables. Therefore it is recommended to take up the bulk of the



slack for the leaf-core trunks around the space near the leaf racks (either in the overhead rails or under the raised floor).



**Figure 1-8:** Cable Bundles between Leaf and Core Switches

## Connect Leaf switches to Core

- Core switches should be generally be connected from the bottom most LIM upwards (starting with LIM 8 up towards LIM 1).
- Core switches should generally be connected from the left-most and right-most ports inward. Ex: 1, 12, 2, 11, 3, 10, ...
- As the core is connected, keep free working space towards the mid-section. Use large tie-wrap strips to keep the core cables evenly curtained between the left and right sides of the rack. All core ports 1-6 for all LIMs should be curtained/tie-wrapped towards the left side of a rack; all core ports 7-12 for all LIMs should be curtained/tie-wrapped towards the right side of a rack.
- After cables are connected to the core switches, take up any excess slack.
- Be mindful of keeping working space around the core switches to ensure ease of future maintenance activities, including replacement of cables and LIMs.

## Connect Nodes to Leaf switches

Take up any excess slack, and tie wrap cables towards the side of the rack if possible and/or appropriate.

## Install Host-side Software

The Topspin host-side drivers will need to be distributed and installed on all nodes of the cluster. Typically this will be done with automated remote copy and remote command execution utilities.



## Install Topspin host-side RPMs, appropriate to the target OS

Installation of the host-side drivers will not include generation of the network scripts to bring up the ib0 interfaces. Refer to the *Topspin HCA/Host Side Drivers Installation Guide* for setting up ib0 initialization scripts.

## Generate /etc/hosts File

The /etc/hosts file should include entries for both the eth0 or ib0 hostnames and addresses. Administrators can decide on using either a central DNS scheme or replicating the /etc/hosts file on all nodes.

## Power on Cluster

In an InfiniBand fabric, the switch running the subnet manager should be powered on last. All subnet managers should be disabled on other switches. The subnet manager should be running on one of the core switches, and after power-on, the standby/backup subnet manager(s) should be enabled with a lower priority. This enables one subnet manager to perform a single, sweep of the entire fabric. If the subnet manager was powered on first, it would perform a new sweep of the fabric every time a new switch came online because of discovered state changes.

## Troubleshoot Bad Cables

An expected failure rate of cables may be in the area of 1% of all cables. For a 500 node non-blocking cluster, this would imply about 10 (out of 1000) bad cables. Identifying and replacing bad cables may prove challenging.

At the CLI for each of the switches, run 'show interface ib all statistics' --- and look for any '**in-unknown-protos**' that are non-zero. These are typically indicative of bad leaf-core cables.

On the host, check '**/proc/topspin/core/ca1/port1/counters**' and look for any non-zero 'Symbol error counter' values. These are typically indicative of bad node-leaf cables.

Once the bad cables have been isolated, it may be easier to simply clip/cut off the connectors from each end. This way, any bad cables can be gently pulled through a large cable bulk fairly friction-free. Any bad cables that cannot be easily pulled back should just be cut/clipped as far up as possible, without impeding/intruding on the main cable mass.

One indicator of system stability is to run 'show logging end' at the primary core CLI, and look for quiesced logging.

If an entire TS-270 line module needs replacement, it may be easier to first remove adjacent and/or higher line modules with cables intact, before replacing the defective LIM. For example, if LIM 3 needs replacement, then first remove LIM 1 (with cables intact) followed by LIM 2 (with cables intact) before replacing LIM 3. This approach may be preferred in situations where workable space has been reduced.

# Installation Risks and Issues

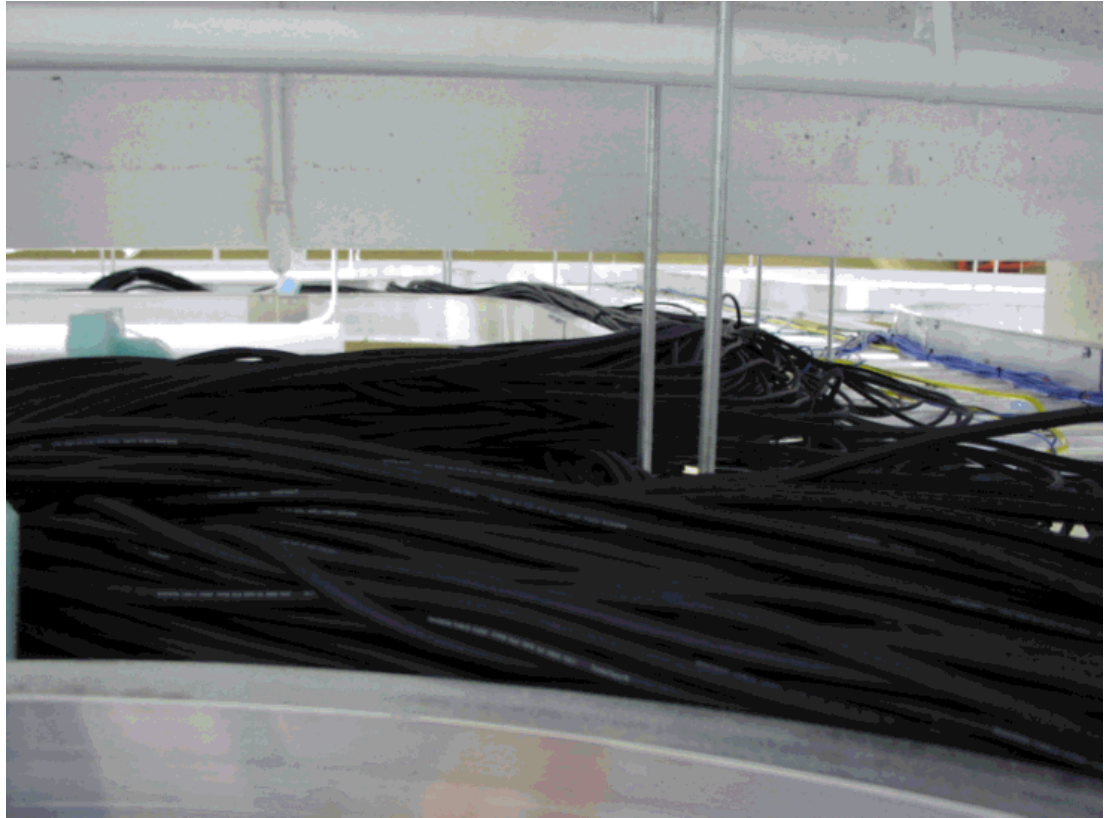
This section addresses the following topics:

- [“IB Core Cable Mass” on page 17](#)
- [“Spares” on page 20](#)

## IB Core Cable Mass

Special attention should be paid to managing the core cable mass, particularly in the following regards.

**Sufficient support should exist in any overhead ceiling trays to support the core cable mass. For the long distance cables used to connect the leaf and core switches, the weight is approximately *1.5 pounds per meter!***



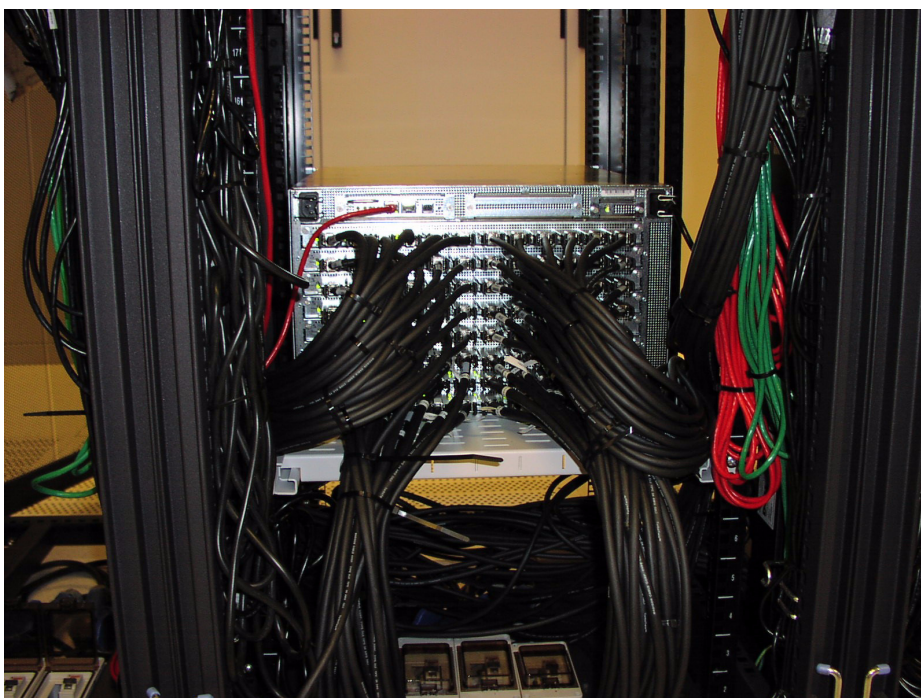
**Figure 1-9: Core Cable Mass**

Sufficient working space should be maintained both within the racks where the core switches reside and in the overhead rails where the core cable mass resides. Free working space tends to be the most precious resource during installation.

If cable bundles are expected to run through the floor (for raised floor installations), it is imperative that the cable bundles do not block airflow used for cooling the racks. This factor may limit the number of switches.



**Figure 1-10:** Side Cable Mass: Bad



**Figure 1-11:** Side Cable Mass: Good

## Spares

Plan in advance to have adequate spare parts available, including spares for:

1. IB cables of all lengths
2. TS-120's and/or TS-120 power supplies
3. TS-270 FRUs
4. HCAs

# Getting Started

---

The following sections appear in this chapter:

- [“Introduction” on page 21](#)
- [“Overview of Set-up Procedures” on page 22](#)
- [“Install HCAs and Drivers in Hosts” on page 22](#)
- [“Install and Power on the InfiniBand Chassis” on page 22](#)
- [“Configure Basic Connectivity” on page 22](#)
- [“Test Network Connectivity” on page 22](#)
- [“Install the GUI \(Element Manager\)” on page 23](#)
- [“Access a Management System” on page 23](#)
- [“Configure Partitions” on page 24](#)
- [“Set User Levels and Passwords” on page 27](#)
- [“\(Optional\) Configure any Ethernet Gateways” on page 28](#)
- [“\(Optional\) Configure any Fibre Channel Gateways” on page 28](#)
- [“\(Optional\) Configure the VFrame Director” on page 28](#)

## Introduction

The sections in this chapter provide configuration information for your Server Switches. Perform these steps on each Server Switch after you complete your hardware installation.

# Overview of Set-up Procedures

Follow the steps below to configure the InfiniBand server switch system.

1. Determine your hardware topology.  
(Optional) If you plan to use the optional Boot Over InfiniBand feature, first refer to the appropriate documentation.
2. Install the Host Channel Adapter and drivers ([page 22](#)).
3. Install and power-on the InfiniBand Chassis ([page 22](#)).
4. Configure Basic Connectivity ([page 22](#)).
5. Test Network Connectivity ([page 22](#)).
6. Install the Element Manager GUI ([page 23](#)).
7. Access a Management System ([page 23](#)).
8. Configure Partitions ([page 24](#)).
9. Set User Level and Access ([page 27](#)).
10. (Optional) Configure the Ethernet Gateway ([page 28](#)).
11. (Optional) Configure the Fibre Channel Gateway ([page 28](#)).
12. (Optional) Configure the VFrame Director ([page 28](#)).

## Install HCAs and Drivers in Hosts

Refer to the *HCA/Host Side Drivers Installation Guide*.

## Install and Power on the InfiniBand Chassis

Refer to the appropriate *Hardware Guide* for installation and power instructions.

## Configure Basic Connectivity

Refer to the appropriate hardware or QuickStart guide for connectivity instructions.

## Test Network Connectivity

After you install network cables, you can verify connectivity by pinging those connected devices from the CLI or pinging between attached hosts.



## Verify Communication Between Hosts

To verify the Topspin device recognizes and successfully links InfiniBand-attached hosts, enter the **ping** command on one host and specify the IP address of another connected host.

```
# ping 10.2.65.50
PING 10.2.0.50 (10.2.0.50) from 10.2.0.41 : 56(84) bytes of data.
64 bytes from 10.2.0.50: icmp_seq=0 ttl=64 time=164 usec
64 bytes from 10.2.0.50: icmp_seq=1 ttl=64 time=144 usec
...
...
6 packets transmitted, 6 packets received, 0% packet loss
round-trip min/avg/max/mdev = 0.135/0.147/0.164/0.017 ms
#
```

## Verify Switch to Host Communication

To verify the InfiniBand device can reach a host on the network, enter either the user-execute or privileged-execute mode on the InfiniBand device, then enter the **ping** command. This is an In-band procedure only.

```
Switch# ping 10.10.253.47
Sending 5 ICMP Echos to 10.10.253.47, 56 data bytes
!!!!!!

Success rate is 100 percent (5/5)
round-trip min/avg/max = 0.000000/0.000000/0.000000 ms
Switch#
```

## Install the GUI (Element Manager)

For details, refer to the *Element Manager User Guide*.

## Access a Management System

Use the following methods to manage the Topspin system:

### CLI Management

Run the Command Line Interface (CLI) from one of the following methods:

- [“Remote Telnet Login” on page 23](#)
- [“Remote SSH Login” on page 24](#)

### Remote Telnet Login

You can Telnet to the Management-Ethernet port on the box from a host on the same network as the Management-Ethernet port, or from any host with a route to the Management-Ethernet network.

To run the CLI remotely:

1. Open a terminal or terminal emulator window. For example: from the command line, enter the **telnet** command with the IP address, or network name, of the Management-Ethernet port.

```
# telnet 10.0.0.47
```

The CLI login prompt (`login:`) is displayed.

2. Enter a CLI user name.

The CLI password prompt (`password:`) is displayed.

3. Enter the CLI user password.

The prompt changes to `Switch-type>` to indicate a successful login. The Topspin system is now ready to receive CLI commands.

## Remote SSH Login

TopspinOS supports SSH2 for secure, encrypted login to the Topspin CLI. SSH is enabled by default, and does not require additional configuration.

1. To login via SSH, use an SSH client (e.g. Putty) to port 22.

## GUI Management

1. Run the Element Manager (GUI) over a TCP/IP network.
2. To login to the GUI, refer to the *Element Manager User Guide*.

## SNMP Management

Any Network Manager running the Simple Network Management Protocol (SNMP) can manage the Topspin system, if the Management Information Base (MIB) is installed correctly. By default, the Topspin GUI is a network manager and uses SNMP v2c as the protocol to communicate between the chassis and the management workstation.

# Configure Partitions

Configure partitions on your IB network.

## Create a Partition ID (P\_Key)

A default partition is configured automatically. The members of a default partition include all connected ports, and provide full membership. However, to create separation between traffic, you must configure specific partitions.

1. Launch Element Manager, if you have not already done so.
2. Select **InfiniBand** --> **Subnet Management**.

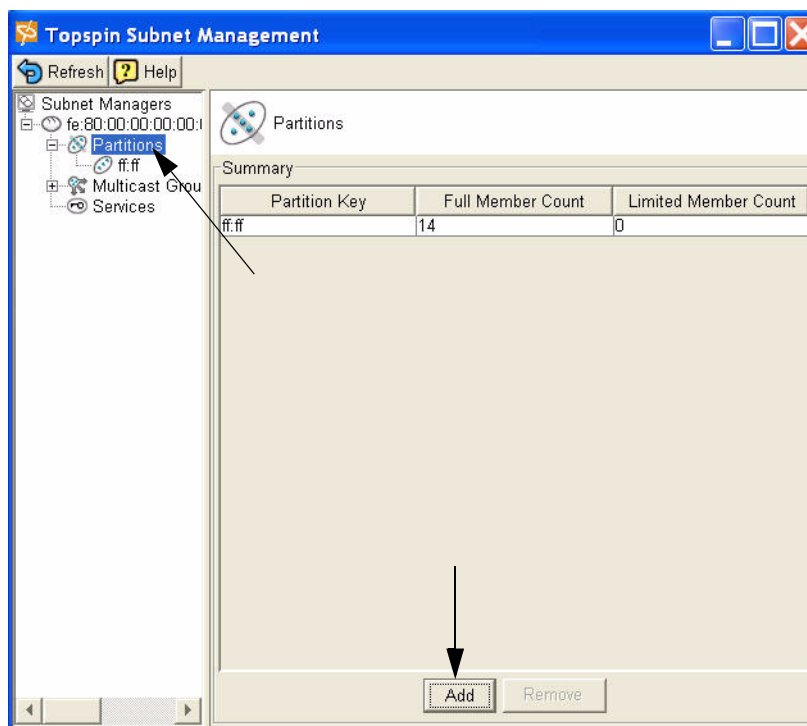
The Subnet Management window appears.

3. Click open the **Subnet Manager** folders in the left window.

The Partitions folder appears.



4. Click on the **Partitions** folder in the left window. The Partitions Summary window appears.



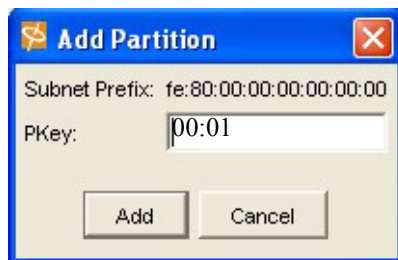
**Figure 2-1: Partitions Summary Window**

5. Click the **Add** button.

The Add Partition dialog box appears.

Enter a Partition key (P\_Key) to identify the new partition.

#### Example



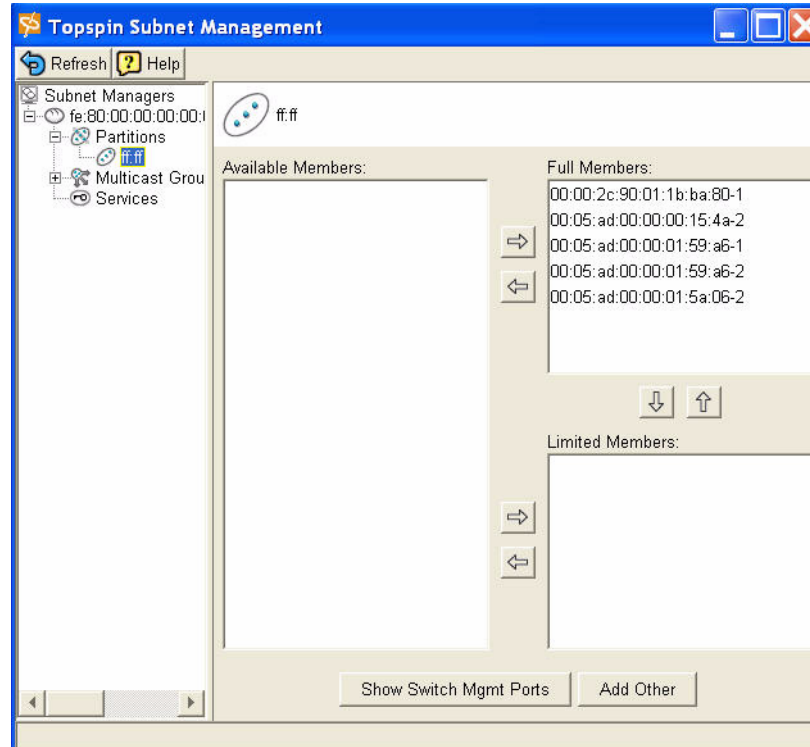
- a. Click the **Add** button.

The new Partition appears in the left window.

## Specify Partition Members and the Membership Type

- b. Click on the new Partition in the left window.

The available partition members appear in the right-side window.



**Figure 2-2: Available Partition Members**

Note that the “Available Members” refers only to members that are known to the Subnet Manager. This includes HCAs and Switches that are already plugged into the fabric as well as manually configured entries.

If you know the GUID and port count of an HCA that has not yet been installed, you can configure it before it is plugged in by using the “Add Other” button.

- c. Click on a member from the Available Member list, and use the arrow button to move it to the Full or Limited member columns.
- d. Click back to the Partitions folder (in the left-side window) when you have selected all of the members for your Partition.

The new Partition appears in the Partition Summary window.

## Maintain Partition Key Information

The configured p\_keys will be needed in completing the configuration of the system.

- e. Configured partition keys must be mapped to any of the following components that exist:
- Host Channel Adapters (HCAs).
  - Ethernet Gateway Bridge-groups.

- Fibre Channel gateways.
- f. If you have multiple InfiniBand switches in your fabric:
  - Exchange the partition configuration between switches by enabling database synchronization, if you have not already done so. This feature is enabled by default.

# Set User Levels and Passwords

## Change Default User Name and Password

For security purposes, since multiple users exist on the system, it is highly recommended that you change the default passwords after initial configuration.

1. Log in to the CLI as a super user. Use the default user name and the default password if they have not already been changed.
2. Enter the privileged-user mode.
3. Enter the global-configuration mode.
4. Enter the **username** command and the **password** keyword to change the user account and user password.

Use the default user name and password if they have not already been changed.

The user name and password are alphanumeric strings of up to 34 characters each.

5. Repeat step 4 to change additional usernames and passwords.

### Example

```
Switch# Login: super
Password: xxxx
Switch> enable
Switch# configure
Switch(config)# username ib-fc_admin password ibFcAdmin
Switch(config)# username ib-fc_admin communitystring
ibFc-commStr
```

6. Exit the global-configuration mode.
7. Use the **show user** command to verify changes.

Only a user with unrestricted privileges may view user information.

```
Switch> show user all
=====
User Information
=====
    username : admin
    password : tospin
    snmp-community : justatest
    permission-level : ib-rw, ip-ethernet-rw, fc-rw
    admin-status : enabled
    num-logins : 0
    num-unsuccessful-logins : 0
    last-login :
    last-unsuccessful-login :

Switch>
```

## **(Optional) Configure any Ethernet Gateways**

Ethernet gateways are activated and enabled automatically upon installation in the chassis.

## **(Optional) Configure any Fibre Channel Gateways**

Fibre Channel gateways are activated and enabled automatically upon installation in the chassis.

Note: If you plan to use the optional VFrame Director, do not create ITLs for any of the Fibre Channel gateways. The VFrame Director will create the ITLs for you, and any previous configurations could create a conflict.

## **(Optional) Configure the VFrame Director**

VFrame Director is an optionally licensed product. If you are using this product to manage your data center, refer to the *VFrame Director Users Guide* and the *Boot Over InfiniBand Configuration Guide*.

# Cable Length Sizing

---

The following sections appear in this chapter:

- [“Sizing” on page 29](#)

## Sizing

[Table 3-1](#) provides cable length sizing information.

**Table 3-1:** Cable Length Sizing

Connection Type	Length	Notes
Within a rack (switch to host usually)	1m	<ul style="list-style-type: none"><li>• Same side connection</li><li>• No obstructions, no cable guides to run through</li></ul>
	3m	<ul style="list-style-type: none"><li>• Any other situation</li></ul>
Between Adjacent Racks	3m	<ul style="list-style-type: none"><li>• Same side connection</li><li>• No obstructions, no cable guides to run through</li><li>• Does not have to span the height of the rack</li></ul>
	5m	<ul style="list-style-type: none"><li>• Same side connection</li><li>• No obstructions, no cable guides to run through</li><li>• Spans the height of the rack</li></ul>
	7m	<ul style="list-style-type: none"><li>• Any other situation</li></ul>

**Table 3-1: Cable Length Sizing (Continued)**

Connection Type	Length	Notes
Between Racks separated by 1 rack in a row, or directly across to another row	5m	<ul style="list-style-type: none"> <li>• Same side connection</li> <li>• No obstructions, no cable guides to run through</li> <li>• Spans the height of the rack</li> </ul>
	7m	<ul style="list-style-type: none"> <li>• Same side connection</li> <li>• No obstructions, no cable guides to run through</li> <li>• Spans the height of the rack</li> </ul>
	10m	<ul style="list-style-type: none"> <li>• Any other situation</li> </ul>
Between Racks separated by 2 racks in a row, or across to another row offset by 1 rack	7m	<ul style="list-style-type: none"> <li>• Same side connection</li> <li>• No obstructions, no cable guides to run through</li> <li>• Spans the height of the rack</li> </ul>
	10m	<ul style="list-style-type: none"> <li>• Same side connection</li> <li>• No obstructions, no cable guides to run through</li> <li>• Spans the height of the rack</li> </ul>
	15m	<ul style="list-style-type: none"> <li>• Any other situation</li> </ul>
Between Racks separated by 3 racks in a row, or across to adjacent row offset by 2 racks, or directly across to a rack 2 rows away	10m	<ul style="list-style-type: none"> <li>• Same side connection</li> <li>• No obstructions, no cable guides to run through</li> <li>• Spans the height of the rack</li> </ul>
	15m	<ul style="list-style-type: none"> <li>• Same side connection</li> <li>• No obstructions, no cable guides to run through</li> <li>• Spans the height of the rack</li> </ul>
Any other situation	15m???	<ul style="list-style-type: none"> <li>• Any other situation will have to be measured to determine whether 15m cable can reach</li> </ul>

# Physical and Environmental Specifications

The following sections appear in this chapter:

- [“Specifications” on page 31](#)

## Specifications

[Table 4-1](#) provides Server Switch dimensions and weights.

**Table 4-1:** Topspin Product Dimension and Weight Table

Product	Dimensions						Max Weight	
	Width		Depth		Height		Pounds	Kilograms
	Inches	mm	Inches	mm	Inches	mm		
TS-120	17	432	22	588.8	1.75	44.45	30	13.6
TS-270	17	432	24	609.8	10.3	262	110	50.1
TS-360	17	432	24	609.8	7	177.8	95	43.1

[Table 4-2](#) provides cable dimensions and weights.

**Table 4-2:** Cable Weights

Cable	Weight
4X IB cable, 1m 28 AWG	0 LB. 5.9 OZ
4X IB cable, 3m 28 AWG	0 LB. 10.8 OZ
4X IB cable, 5m 26 AWG	1 LB. 3.8 OZ
4X IB cable, 7m 24 AWG	2 LBS. 2 OZ (extrapolated from 10m with same connector weighs)
4X IB cable, 10m 24 AWG	2 LBS. 15 OZ

**Table 4-2: Cable Weights (Continued)**

Cable	Weight
4X IB cable, 15m 24 AWG	3 LBS. 12 OZ

Table 4-3 provides product operating environments

This reflects the cable only, and does not include packaging.

**Table 4-3: Topspin Product Operating Environment**

Product	Temperature		Humidity	
	Operating	Non-Operating	Operating	Non-Operating
TS-120	0 ~ 45° Celsius	-40 ~ 70° Celsius	20% ~ 80%	5% ~ 90%
	0 ~ 42° Celsius	-40 ~ 70° Celsius	20% ~ 80%	5% ~ 90%
	0 ~ 35° Celsius	-25 ~ 70° Celsius	20% ~ 80%	5% ~ 95%

Table 4-4 displays power and cooling requirements.

**Table 4-4: Topspin Product Power and Cooling Requirements**

Product	Max Power Consumption (W)	Max Calorific Power (kJ/h)	Air Volume (m3/min)	Capacity of Electric Power (A @ 120V)
TS-120	65	234	1.4	0.54
TS-270	600	2880	4.7	5.00
TS-360	650	2340	4.1	5.42
Fibre Channel Gateway (2 Ports)	25	90	N/A	0.21
Ethernet Gateway (6 Ports)	26	93.6	N/A	0.22



## P-Touch Editor Notes

---

The following sections appear in this chapter:

- [“Introduction” on page 33](#)

### Introduction

- There are two modes of operation: database mgmt and label formatting.
- Database mgmt allows the importing of Excel CSV files by way of
- “File->New->Database” or “File->New->Database (for importing)”.
- “File->New->Label” allows for the formatting of a new label. Make sure the label size in “Printer->Page Setup” matches for the size of the cartridge being used.
- Create a mirrored format for the label, so that the label can be read from any orientation. Example “node 385 leaf 33/13    node 385 leaf 33/13”.
- To get the database linked in to the label format, highlight the label format window, and go “Insert->Database Field”.
- For printing:
  - Go to the window that displays the database under operation and select/highlight the set of rows to be printed.
  - Click back to the label-format window (which changes the program menu format).
  - File->Print
  - Click the button for “Print Range: Selected Records”. Number of copies should be 2 (one for each end of the cable).



## IB Port Test

---

The following sections appear in this chapter:

- [“Introduction” on page 35](#)
- [“Script” on page 35](#)

## Introduction

Run the following script to verify your cable connectivity.

## Script

```
#!/usr/bin/expect --

#####
#####
#
# ib_port_test -- Expect script to check/report on IB ports in a
# chassis.
#
# $Id: ib_port_test,v 1.4 2004/11/01 22:38:01 scott Exp $
#
#####
#####
#
#                               (c) COPYRIGHT 2001, 2002, 2003, 2004
#                               Topspin Communications, Inc.
#                               Mountain View, CA
#####
#####

puts "----- Performing InfiniBand Port Test -----"
```

```

# Check the number of arguments and if needed print the usage
set argc [llength $argv]
if { $argc == 0 } {
    puts "Error: Arguments given not sufficient\n"
    puts "Usage: $argv0 <ipaddr> \[ <user> <password> \] \n"
    exit 0
}

set ipAddr [lindex $argv 0]

if { $argc == 3 } {
    set user [lindex $argv 1]
    set password [lindex $argv 2]
} else {
    # Use default user name and password
    set user "super"
    set password "super"
}

# Lists to process information from "show interface ib all"
set portList {}
set adminSpeedList {}
set operSpeedList {}
set portSpeedMismatchList {}

# Lists to process information from "show interface ib all statistics"
set portListStats {}
set inDiscardsList {}
set inErrorsList {}
set inUnknownProtosList {}
set outDiscardsList {}
set outErrorsList {}
set portStatErrorList {}

# TODO: Remove the IBM BladeCenter related customization in the prompt
(?)
# TODO: Port this script to work for other OEMs also (?)
# Pattern for CLI prompt
set prompt "((Topspin-|IBM BladeCenter)\[A-Z0-9\]*(>|#))"

# Turn off expect output
log_user 0

# Login into the chassis
spawn telnet $ipAddr

expect_after {
    timeout {
        puts "expect_after: Timeout occurred."
        exit 0
    }
    eof {
        puts "expect_after: Unexpected end of file."
        exit 0
    }
}

expect {
    -re "Login|IBM BladeCenter login" {
        send "$user\r"
        expect "Password:"
    }
}

```

```

        send "$password\r"
        expect -re $prompt
        send "enable\r"
        expect -re $prompt
        # Turn off paging
        send "terminal length 0\r"
        expect -re $prompt
        send "show version\r"
        expect -re $prompt
        # Gather info about IB ports and their speeds
        send "show interface ib all\r"
        # Parse output one line at a time.
        expect {
        -re "([^\r]*)\r+\n" {
            #puts "GOT LINE: $expect_out(1,string)"
            if { [regexp "(port : \[0-9]+\[/]*\[0-9]*)\$"
$expect_out(1,string) ignore newPort] } {
                #puts "new port $newPort"
                lappend portList $newPort
            }
            if { [regexp "(admin-speed : \[\.a-zA-Z0-9]+\)\$"
$expect_out(1,string) ignore newAdminSpeed] } {
                #puts "new adminSpeed $newAdminSpeed"
                lappend adminSpeedList $newAdminSpeed
            }
            if { [regexp "(oper-speed : \[\.a-zA-Z0-9]+\)\$"
$expect_out(1,string) ignore newOperSpeed] } {
                #puts "new operSpeed $newOperSpeed"
                lappend operSpeedList $newOperSpeed
            }
        }
        exp_continue
    }

    -re $prompt { }
}
send "\r"
expect -re $prompt

# TODO: rewrite to use fewer lists

# Gather info about IB port errors
send "show interface ib all statistics\r"
expect {
    -re "([^\r]*)\r+\n" {
        #puts "GOT LINE: $expect_out(1,string)"
        if { [regexp "(port : \[0-9]+\[/]*\[0-9]*)\$"
$expect_out(1,string) ignore newPort] } {
            #puts "new port $newPort"
            lappend portListStats $newPort
        }
        if { [regexp "(in-discards : \[\\-0-9]+\)\$" $expect_out(1,string)
ignore newStat] } {
            lappend inDiscardsList $newStat
        }
        if { [regexp "(in-errors : \[\\-0-9]+\)\$" $expect_out(1,string)
ignore newStat] } {
            lappend inErrorsList $newStat
        }
        if { [regexp "(in-unknown-protos : \[\\-0-9]+\)\$"
$expect_out(1,string) ignore newStat] } {
            lappend inUnknownProtosList $newStat
        }
    }
}

```

```

        if { [regex "(out-discards : \[-0-9+\)\$"
$expect_out(1,string) ignore newStat] } {
        lappend outDiscardsList $newStat
        }
        if { [regex "(out-errors : \[-0-9+\)\$" $expect_out(1,string)
ignore newStat] } {
        lappend outErrorsList $newStat
        }
        if { [regex "(out-unknown-protos : \[-0-9+\)\$"
$expect_out(1,string) ignore newStat] } {
        lappend outUnknownProtosList $newStat
        }

        exp_continue
    }

    -re $prompt { }
    }
    send "\r"
    expect -re $prompt
}

timeout {
    puts "Error: Timeout trying to login into the chassis"
    exit 0
}
}

# Process the info from "show interface ib all" output
# to find any port speed mismatches
set count 0
set portSpeedMismatchFlag 0
foreach i $portList {
    set adminSpeedVar [lindex $adminSpeedList $count]
    split $adminSpeedVar
    set adminSpeed [lindex $adminSpeedVar 2]
    set operSpeedVar [lindex $operSpeedList $count]
    split $operSpeedVar
    set operSpeed [lindex $operSpeedVar 2]
    #puts "port $i: adminSpeed $adminSpeed, operSpeed $operSpeed"
    if { $adminSpeed != $operSpeed && $operSpeed != "unknown" } {
        set portSpeedMismatchFlag 1
        lappend portSpeedMismatchList "Port speed for $i is $operSpeed
instead of $adminSpeed"
    }
    set count [expr $count+1]
}

# Process the info from "show interface ib all statistics"
# to find any port errors like in-discards, in-errors,
# in-unknown-protos, out-discards and out-errors.
set count 0
set portStatErrorCountFlag 0
foreach i $portListStats {

    # For in-discards
    set inDiscardsVar [lindex $inDiscardsList $count]
    split $inDiscardsVar
    set inDiscardsCount [lindex $inDiscardsVar 2]
    if { $inDiscardsCount != 0 } {
        set portStatErrorCountFlag 1
    }

```

```

        lappend portStatErrorList "in-discards for $i is
$inDiscardsCount"
    }

    # For in-errors
    set inErrorsVar [lindex $inErrorsList $count]
    split $inErrorsVar
    set inErrorsCount [lindex $inErrorsVar 2]
    if { $inErrorsCount != 0 } {
        set portStatErrorCountFlag 1
        lappend portStatErrorList "in-errors for $i is $inErrorsCount"
    }

    # For in-unknown-protos
    set inUnknownProtosVar [lindex $inUnknownProtosList $count]
    split $inUnknownProtosVar
    set inUnknownProtosCount [lindex $inUnknownProtosVar 2]
    if { $inUnknownProtosCount != 0 } {
        set portStatErrorCountFlag 1
        lappend portStatErrorList "in-unknown-protos for $i is
$inUnknownProtosCount"
    }

    # For out-discards
    set outDiscardsListVar [lindex $outDiscardsList $count]
    split $outDiscardsListVar
    set outDiscardsCount [lindex $outDiscardsListVar 2]
    if { $outDiscardsCount != 0 } {
        set portStatErrorCountFlag 1
        lappend portStatErrorList "out-discards for $i is
$outDiscardsCount"
    }

    # For out-errors
    set outErrorsListVar [lindex $outErrorsList $count]
    split $outErrorsListVar
    set outErrorsCount [lindex $outErrorsListVar 2]
    if { $outErrorsCount != 0 } {
        set portStatErrorCountFlag 1
        lappend portStatErrorList "out-errors for $i is
$outErrorsCount"
    }
    set count [expr $count+1]
}

# Print the results to the user.
if { $portSpeedMismatchFlag == 0 } {
    puts "Check IB Port Speed Mismatches ..... PASS"
} else {
    puts "Check IB Port Speed Mismatches ..... FAIL"
    puts "    Reason: found port speed mismatch for the following:"
    foreach j $portSpeedMismatchList {
        puts "        $j"
    }
}

if { $portStatErrorCountFlag == 0 } {
    puts "Check IB Port Errors ..... PASS"
} else {
    puts "Check IB Port Errors ..... FAIL"
    puts "    Reason: found non-zero entries for the following:"
    foreach k $portStatErrorList {

```

```
        puts "      $k"
    }
}

puts "----- Done -----"
```



# Index

## C

### Connectivity

basic .....	22
establishing .....	22

## N

### Network Connectivity

testing .....	22
---------------	----

## R

Remote SSH Login .....	24
Remote Telnet Login .....	23

## S

set-up procedures .....	22
-------------------------	----