Sun Fire™ Link System Overview

Sun Microsystems, Inc.
4150 Network Circle
Santa Clara, CA 95054 U.S.A.
650-960-1300

Part No. 816-0697-11
May 2003, Revision A

Send comments about this document to: docfeedback@sun.com

Copyright 2003 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 embodied in the product 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, AnswerBook2, docs.sun.com, Java, Java Dynamic Management Kit, Solaris, Sun Fire, Sun HPC ClusterTools, and RSM are trademarks or registered trademarks of Sun Microsystems, Inc. in the U.S. and in other countries.

All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. in the U.S. and in other countries. Products bearing SPARC trademarks are based upon an architecture developed by Sun Microsystems, Inc.

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.

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 2003 Sun Microsystems, Inc., 4150 Network Circle, Santa Clara, California 95054, Etats-Unis. Tous droits réservés.

Sun Microsystems, Inc. a les droits de propriété intellectuels relatants à la technologie incorporée dans le produit qui est décrit dans ce document. En particulier, et sans la limitation, ces droits de propriété intellectuels peuvent inclure un ou plus des brevets américains énumérés à http: / / www.sun.com / patents et un ou les brevets plus supplémentaires ou les applications de brevet en attente dans les Etats-Unis et dans les autres pays.

Ce produit ou document est protégé par un copyright et distribué avec des licences qui 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, parquelque moyen que ce soit, sans l'autorisation préalable et écrite de Sun et de ses bailleurs de licence, s'il y ena.

Le logiciel détenu par des tiers, et qui comprend la technologie relative aux polices de caractères, est protégé par un copyright et licencié par des fournisseurs de Sun.

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

Sun, Sun Microsystems, the Sun logo, AnswerBook2, docs.sun.com, Java, Java Dynamic Management Kit, Solaris, Sun Fire, Sun HPC ClusterTools, et RSM sont des marques de fabrique ou des marques déposées de Sun Microsystems, Inc. aux Etats-Unis et dans d'autres pays.

Toutes les marques SPARC sont utilisées sous licence et sont des marques de fabrique ou des marques déposées de SPARC International, Inc. aux Etats-Unis et dans d'autres pays. Les produits protant les marques SPARC sont basés sur une architecture développée par Sun Microsystems, Inc.

L'interface d'utilisation 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 pour la recherche et le développment du concept des interfaces d'utilisation visuelle ou graphique pour l'industrie de l'informatique. Sun détient une license non exclusive do Xerox sur l'interface d'utilisation graphique Xerox, cette licence couvrant également les licenciées de Sun qui mettent en place l'interface d 'utilisation graphique OPEN LOOK et qui en outre se conforment aux licences écrites de Sun.

LA DOCUMENTATION EST FOURNIE "EN L'ÉTAT" ET TOUTES AUTRES CONDITIONS, DECLARATIONS ET GARANTIES EXPRESSES OU TACITES SONT FORMELLEMENT EXCLUES, DANS LA MESURE AUTORISEE PAR LA LOI APPLICABLE, Y COMPRIS NOTAMMENT TOUTE GARANTIE IMPLICITE RELATIVE A LA QUALITE MARCHANDE, A L'APTITUDE A UNE UTILISATION PARTICULIERE OU A L'ABSENCE DE CONTREFAÇON.

Contents

Preface v

1. Introduction 1

Sun Fire Link Clusters 2

Hardware Striping 3

Managing Sun Fire Link Clusters 4

2. Overview of Sun Fire Link Cluster Configurations 5

Direct-Connect Configurations 6

Switched Configurations 7

Sun Fire Link Partitions 9

Striping 9

3. Overview of Sun Fire Link Hardware 11

Sun Fire Link Assembly 11

Sun Fire Link Assembly for Sun Fire 6800 systems 11

Sun Fire Link Assembly for Sun Fire 15K / 12K Systems 14

Sun Fire Link Optical Module 16

Sun Fire Link Optical Cables 16

Sun Fire Link Switch 16

4. Overview of Sun Fire Link Software 19

Sun Fire Link Cluster Driver Software 19

Sun Fire Link Switch Software 19

Sun Fire Link Administration Software 20

Sun Fire Link Manager 20

Sun Management Center / FM Console 21

FM Proxies 21

Sun Management Center Server 21

Sun Fire Link Device Drivers 22

SNMP Agents 22

Sun Management Center Agent / System Controller Proxy, Switch Proxy 22

Index 23

Preface

This document introduces the Sun Fire™ Link interconnect technology, providing an overview of the system's features, as well as its main hardware and software components. This manual also describes the basic Sun Fire Link network configurations that can be achieved with the Sun Fire 6800 and Sun Fire 15K/12K servers from Sun Microsystems, Inc.™

How This Book Is Organized

Chapter 1 introduces the principal architectural concepts underlying the Sun Fire Link interconnect.

Chapter 2 describes the rules governing Sun Fire Link network configurations.

Chapter 3 provides an overview of the main hardware components in a Sun Fire Link system.

Chapter 4 provides an overview of the main software components in a Sun Fire Link system.

Using UNIX Commands

This document might not contain information on basic UNIX® commands and procedures such as shutting down the system, booting the system, and configuring devices.

See one or more of the following for this information:

- *Solaris Handbook for Sun Peripherals*

- Online documentation for the Solaris™ operating environment
- Other software documentation that you received with your system

Typographic Conventions

Typeface*	Meaning	Examples
`AaBbCc123`	The names of commands, files, and directories; on-screen computer output	Edit your `.login` file. Use `ls -a` to list all files. `% You have mail.`
`AaBbCc123`	What you type, when contrasted with on-screen computer output	`% su` `Password:`
AaBbCc123	Book titles, new words or terms, words to be emphasized. Replace command-line variables with real names or values.	Read Chapter 6 in the *User's Guide*. These are called *class* options. You *must* be superuser to do this. To delete a file, type `rm` *filename*.

* The settings on your browser might differ from these settings.

Shell Prompts

Shell	Prompt
C shell	*machine-name*%
C shell superuser	*machine-name*#
Bourne shell and Korn shell	$
Bourne shell and Korn shell superuser	#

Related Documentation

Application	Title	Part Number
Task map	*Task Map for Getting* Started With a Sun Fire Link Network	816-0041
Hardware installation	*Sun Fire Link Hardware Installation Guide*	806-1396
Software installation (with Sun MC 3.0)	*Sun Fire Link Software Installation Guide*	806-1401
Software installation (with Sun MC 3.5)	*Sun Fire Link Software Installation Guide*	817-0401
Administration	*Sun Fire Link Fabric Administrator's Guide*	806-1405
Service	Sun Fire Link Service Manual	805-7363
Switch	*Sun Fire Link Switch Installation and Service Manual*	806-1397

Accessing Sun Documentation

You can view, print, or purchase a broad selection of Sun™ documentation, including localized versions, at:

`http://www.sun.com/documentation`

Sun Welcomes Your Comments

Sun is interested in improving its documentation and welcomes your comments and suggestions. You can email your comments to Sun at:

`docfeedback@sun.com`

Please include the part number (816-0697-10) of your document in the subject line of your email.

CHAPTER 1

Introduction

Sun Fire Link is a high-bandwidth, low-latency cluster interconnect used to expand Sun Fire 6800 and 15K/12K system capabilities beyond the physical boundaries of an individual system chassis. See FIGURE 1-1.

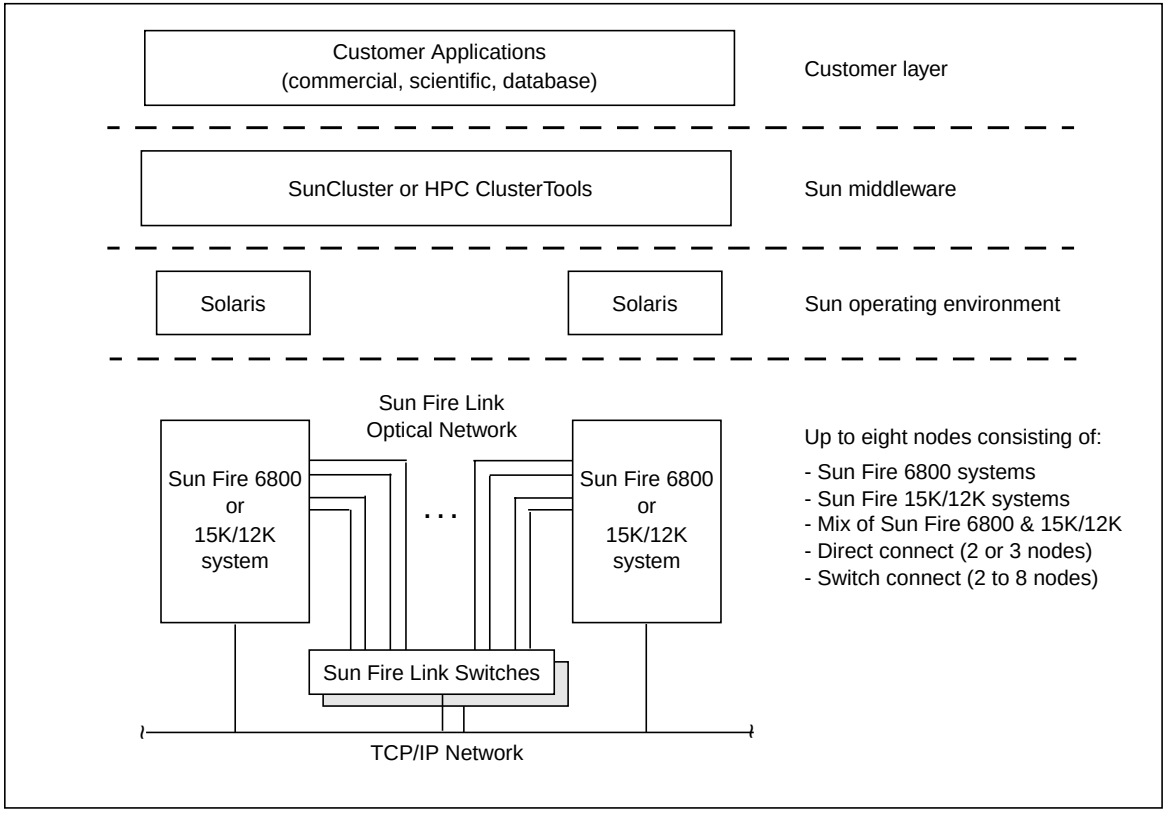

FIGURE 1-1 Sun Fire Link System View

Sun Fire Link Clusters

A Sun Fire Link cluster consists of up to eight Sun Fire 6800 and/or Sun Fire 15K/12K nodes connected to each other by a Sun Fire Link optical network. Each node runs a separate instance of Solaris under a layer of clustering software, which can be either Sun™ Cluster or Sun HPC ClusterTools™ software. For some configurations, the interconnect hardware will include Sun Fire Link switches as well.

A Sun Fire Link cluster also includes a TCP/IP network for cluster administration traffic. This network connects all cluster components that exchange control and status/error information.

Both Sun Cluster and Sun HPC ClusterTools make use of the Remote Shared Memory (RSM™) interface for internode communication across a Sun Fire Link network. RSM is a Sun messaging interface that is highly efficient for remote memory operations.

For Sun Fire Link clusters of two or three nodes, the network connections can be point-to-point (direct-connect topology) or through Sun Fire Link switches. For larger clusters (four to eight nodes), Sun Fire Link switches are required.

The system's interface to the Sun Fire Link network is provided by a Sun Fire Link-specific I/O subsystem called the Sun Fire Link assembly. These are installed in standard system I/O slots. Each Sun Fire Link assembly contains two optical transceiver modules called Sun Fire Link optical modules. Each optical module supports a full-duplex optical link.

The Sun Fire Link assemblies are installed in pairs to enhance availability and to support message striping for higher bandwidth. Each Sun Fire Link assembly pair is in one of the system's compute domains. This means a compute domain that is part of a Sun Fire Link cluster contains four optical link connections to the Sun Fire Link network. See FIGURE 1-2

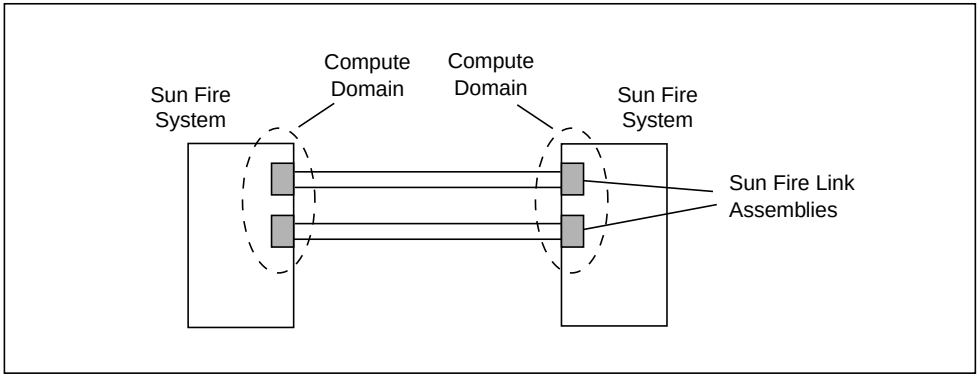

FIGURE 1-2 Compute Domains Connected by Four Optical Links

Note – A compute domain that is connected to a Sun Fire Link network is said to be *clustered*.

The Sun Fire Link switch is an eight-port optical switch housed in a 19-inch rackmount chassis. A switch contains eight optical transceiver ports of the same type as are used in the Sun Fire Link assembly. Each switch port handles one optical network link.

See Chapter 3 for additional information about Sun Fire Link hardware components.

Hardware Striping

The pairing of optical links in each Sun Fire Link assembly and the pairing of Sun Fire Link assemblies in a compute domain means message traffic for that domain can be striped across two or four optical links for higher bandwidth. This link redundancy also provides protection against link failures. If a link goes down, an alternate path is automatically provided for the message traffic the failed link had been handling. See Chapter 2 for additional information on hardware striping in a Sun Fire Link network.

Managing Sun Fire Link Clusters

From an administrative perspective, the nodes and switches in a Sun Fire Link cluster constitute a pool of network resources, called a *fabric*, which can be configured to form one or more RSM-based networks, called *partitions*. The Sun Fire Link resources contained in a fabric are managed by a Sun Fire Link software module called the *Sun Fire Link Manager* (FM).

Multiple fabrics can be created from a set of interconnected Sun Fire systems, with each fabric being administered by a separate FM instance. All Sun Fire Link cluster resources within a Sun Fire system chassis are contained in one fabric and managed by the same FM, thereby protecting against configuration conflicts when creating Sun Fire Link partitions.

The Sun Fire Link Manager and an associated Sun Fire Link Manager Console are installed on a Solaris host called the *Management Station*. The Management Station is not part of the Sun Fire Link cluster. Instead, the Sun Fire Link Manager and its console communicate with the fabric components over the TCP/IP network.

If the Sun Management Center software is used to administer the cluster, its graphical interface can be used for administrative control over Sun Fire Link partitions, such as bringing optical links up and down, adding and removing nodes from a partition, and specifying a partition's link topology. The graphical interface can also be used to monitor the status of various conditions in the partition. The Sun Fire Link Manager also supports a command-line interface, which can be used in place of the Sun Management Center interface for most cluster administration tasks.

Note – It is strongly recommended that the Sun Management Center software be used as the principal tool for administering Sun Fire Link clusters. The Sun Management Center facilities offer superior capabilities for every aspect of configuring, monitoring, and managing of Sun Fire Link network resources.

Chapter 2 and Chapter 4 provide additional information on cluster configuration and cluster management topics. For a more detailed information on cluster configurations, see the *Sun Fire Link System Service Manual*. For detailed instructions on administering Sun Fire Link clusters, see the *Sun Fire Link Fabric Administrator's Guide*.

CHAPTER 2

Overview of Sun Fire Link Cluster Configurations

Sun Fire Link assemblies are installed pairs in place of the standard Sun Fire 6800 or Sun Fire 15K/12K I/O assembly. Each Sun Fire Link assembly pair is part of a Sun Fire 6800 compute domain. Depending on the type of topology used—direct connect or switched—up to eight domains can be connected to a Sun Fire Link network, as follows:

- *Direct connect* – 2 or 3 domains
- *Switched* – 2 to 8 domains

Note – A Sun Fire Link cluster can contain a mix of Sun Fire 6800 and Sun Fire 15K/12K servers in any combination that satisfies the Sun Fire Link domain limits.

A Sun Fire 6800 server can have one domain connected to the Sun Fire Link optical network. This means the server will contain two Sun Fire Link assemblies, each of which will have two optical link connections to the network.

In a Sun Fire 15K/12K server, up to four domains can contain Sun Fire Link assembly pairs. Again, each domain would have four optical links, for a maximum count of 16 optical links for the server as a whole. This is summarized in the following chart:

Node Type	Domains with Sun Fire Link Assemblies	Total Number of Optical Links per Server
Sun Fire 6800	1	4
Sun Fire 15K/12K	1	4
	2	8
	3	12
	4	16

Direct-Connect Configurations

When a Sun Fire Link network does not include switches, both ends of the optical cables connect directly to the server chassis. FIGURE 2-1 illustrates two- and three-node direct-connect configurations.

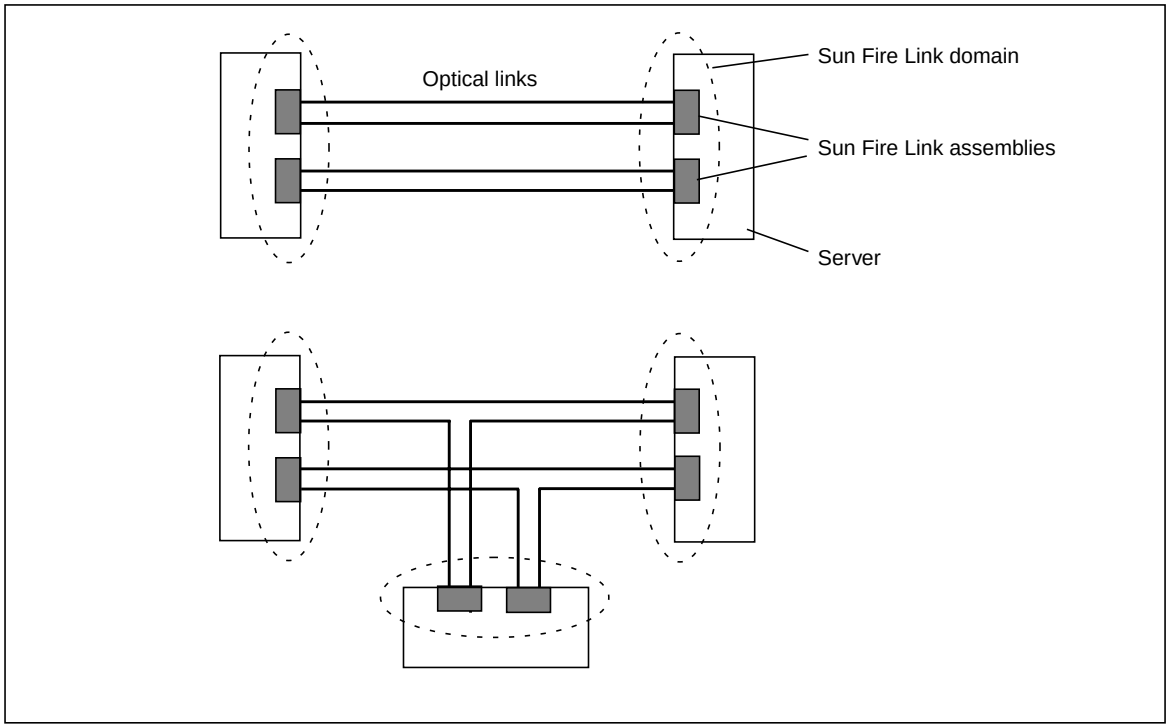

FIGURE 2-1 Two- and Three-Node Direct-Connect Configurations

Switched Configurations

There are two standard switch-based Sun Fire Link network configurations. One configuration contains up to four domains and two Sun Fire Link switches. The other configuration contains up to eight domains and four Sun Fire Link switches. These configurations are illustrated in FIGURE 2-2 and FIGURE 2-3.

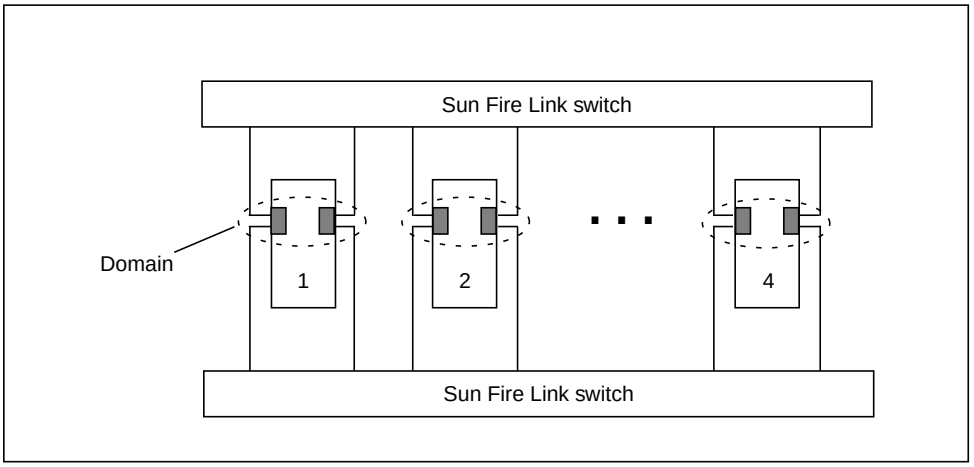

FIGURE 2-2 Two- to Four-Node, Two-Switch Configurations

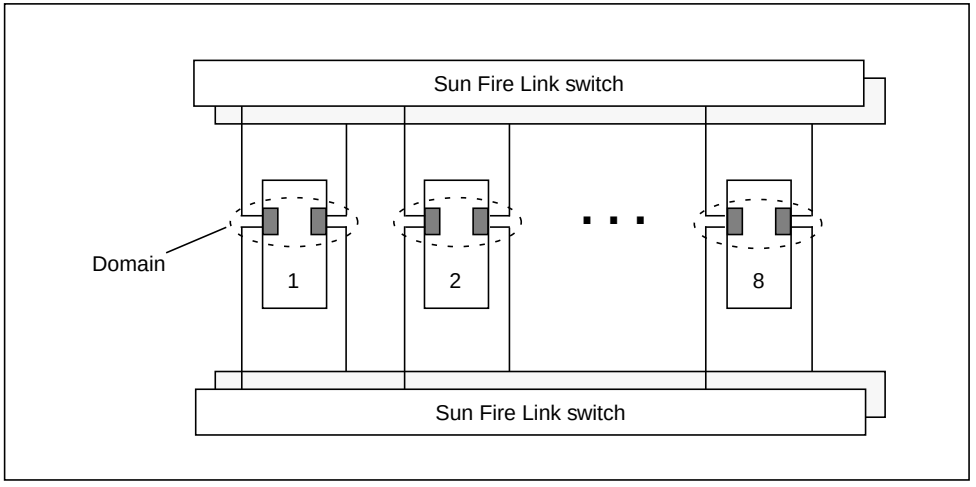

FIGURE 2-3 Five- to Eight-Node, Four-Switch Configurations

Sun Fire Link Partitions

A Sun Fire Link partition is a group of clustered domains that share a network of logical communication paths. That is, the Sun Fire Link Manager creates a partition by mapping the link interfaces of all the domains in the partition as end points in a set of paths that connect all the domains. If the network includes switches, the map includes switch end point information as well.

Striping

When a Sun Fire Link partition is created, the topology of the network paths connecting the domains is specified. The possible choices are:

- No striping – All links are available to carry independent message traffic. This topology offers maximum redundancy for link failure protection.
- Level-two striping – The two Sun Fire Link assemblies in a domain can be allocated to different partitions, with each assembly being managed by a separate software controller. This *dual-controller* configuration is used in Sun Fire Link clusters that are running Sun Cluster applications—that is, applications for which failover support is critical. If the Sun Fire Link interface in the primary controller / partition fails, the job automatically switches to the secondary controller / partition, allowing it to continue execution. Since the Sun Fire Link assembly has two optical network ports, message striping in a dual-controller configuration is referred to as level-two, or two-way striping.
- Level-four striping – In clusters that have Sun HPC ClusterTools software running, both Sun Fire Link assemblies in a domain are configured into the same partition and are managed by a single software controller. In this configuration, messages can be striped across all four of the domain's optical ports, resulting in maximum bandwidth. If any of the four links should fail, the application continues to operate, but at a reduced message-passing bandwidth.

Because both Sun Fire 6800 and Sun Fire 15K/12K domains have the same Sun Fire Link assembly and optical link redundancy, the message striping and failover protection features are the same for both platforms. There is a restriction, however, for the three-node direct-connect configuration. See TABLE 2-1 for a summary of the striping possibilities:

TABLE 2-1 Summary of Link Striping Capabilities

Domains in Network	Connection Type	Possible Striping Levels
2	Direct	2 or 4
3	Direct	2
2-8	Switched	2 or 4

Three domains in a direct-connect configuration are limited to two-way link striping because each domain must be connected to two other domains (FIGURE 2-1). But, if a three-domain network includes switches, message routing is virtual rather than hardwired, so the domains are not restricted in their path options and can use four-way striping.

CHAPTER 3

Overview of Sun Fire Link Hardware

This chapter provides a high-level description of the principal hardware assemblies and components contained in a Sun Fire Link network. See the *Sun Fire Link Hardware Reference Manual* for more information.

Sun Fire Link Assembly

There are two versions of the Sun Fire Link assembly, one for Sun Fire 6800 systems and the other for Sun Fire 15K/12K systems. The differences between the two Sun Fire Link assembly types reflect the different chassis architectures of the two server models.

Note – Both versions of the Sun Fire Link assembly occupy standard chassis I/O slots in their respective Sun Fire systems—that is, each Sun Fire Link assembly is installed in place of a standard Sun Fire 6800 or Sun Fire 15K/12K I/O assembly.

The chief characteristics of the two Sun Fire Link assembly versions, including their differences, are described separately in these sections:

- "Sun Fire Link Assembly for Sun Fire 6800 systems" on page 11
- "Sun Fire Link Assembly for Sun Fire 15K/12K Systems" on page 14

Sun Fire Link Assembly for Sun Fire 6800 systems

FIGURE 3-1 shows block diagram and a mechanical view of the Sun Fire Link assembly for Sun Fire 6800 systems.

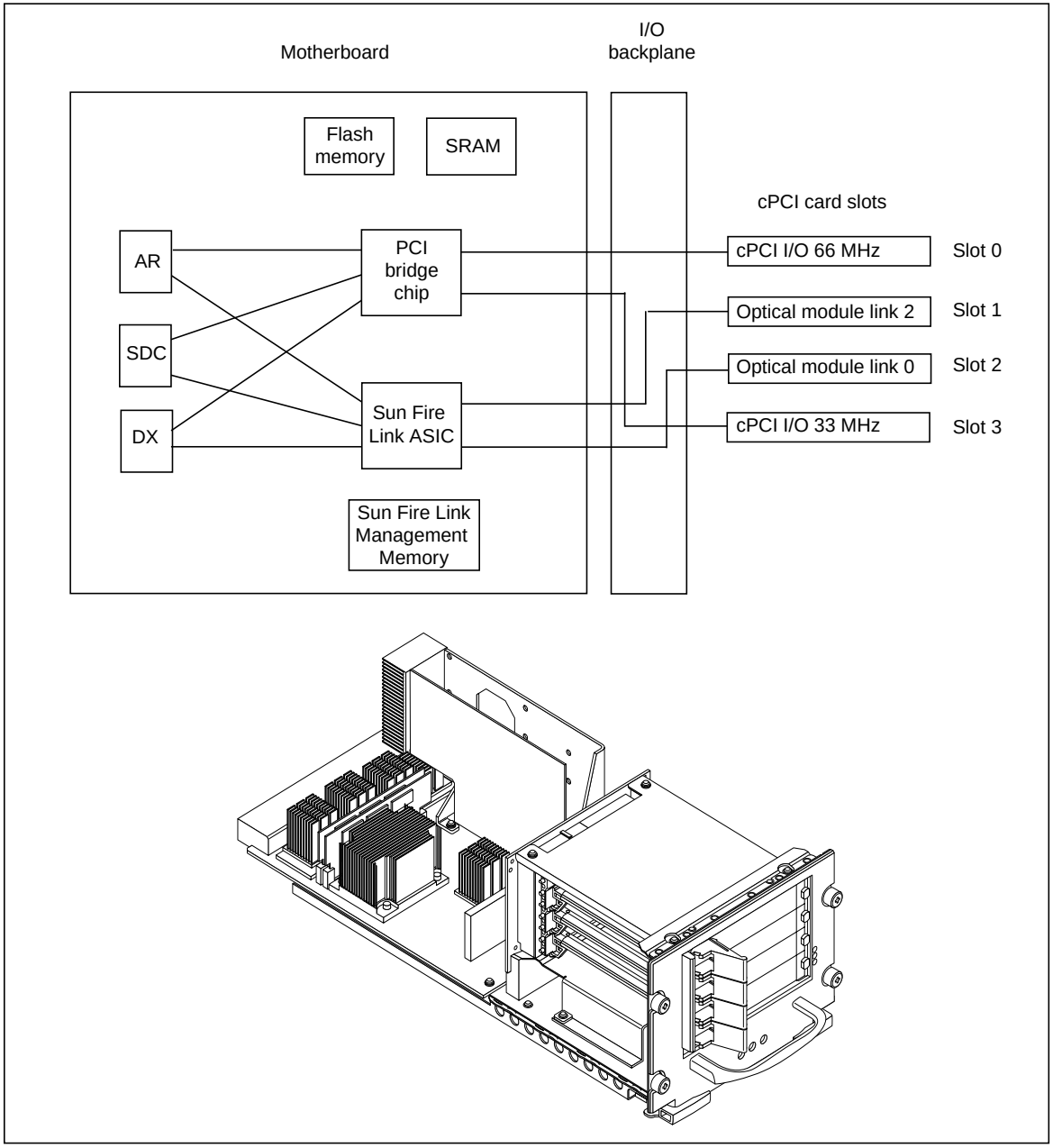

FIGURE 3-1 Sun Fire Link Assembly for Sun Fire 6800 Systems

The Sun Fire Link assembly for Sun Fire 6800 systems contains a four-slot, compact PCI (cPCI) card cage, with the slots allocated as follows:

- Slot 0 – Available for a standard cPCI I/O card; clocked at 66 MHz.
- Slot 1 – Sun Fire Link optical module for link 2.
- Slot 2 – Sun Fire Link optical module for link 0.
- Slot 3 – Available for a standard cPCI I/O card; clocked at 33 MHz.

Note – The Sun Fire Link optical modules are implemented as cPCI form factor cards, but are cPCI compliant only from a mechanical and form factor perspective. The modules are not cPCI compliant with respect to electrical and signalling characteristics.

The Sun Fire Link assembly motherboard includes a PCI bridge chip, which serves as the interface between the cPCI adapter cards (slots 0 and 3) and the system's system interconnect. This same cPCI interface is used in the Sun Fire 6800 I/O assembly.

A separate chip, called the *Sun Fire Link ASIC*, provides the interface between the Sun Fire Link optical modules and the system interconnect. The Sun Fire Link ASIC is a high-performance, network management chip that supports cluster protocols. A single Sun Fire Link ASIC handles both optical links 0 and 2 for the Sun Fire Link assembly on which it resides. The Sun Fire Link ASIC also supports multi-hop routing, retransmission of corrupted packets, and resynchronization of connected nodes.

A Sun Fire 6800 system contains two Sun Fire Link assemblies, for a total of four optical links, all in a single domain. These assemblies must always occupy chassis I/O slots IB8 and IB9. FIGURE 3-2 shows a pair of Sun Fire Link assemblies installed in a Sun Fire 6800 chassis. Note that the two assemblies have a mirror-image orientation with respect to one another. This means the following slot relationships apply to both IB8 and IB9:

- cPCI slot 0 (66 MHz adapter) is closest to the chassis centerline.
- Optical link 2 is the second slot away from the centerline.
- Optical link 0 is the third slot away from the centerline.
- cPCI slot 3 (33 MHz adapter) is farthest from the chassis centerline.

The assembly in IB9 is inverted with respect to IB8 so that the transmit and receive connectors on the optical links are in opposite positions—transmit is on top in IB8 and on the bottom in IB9.

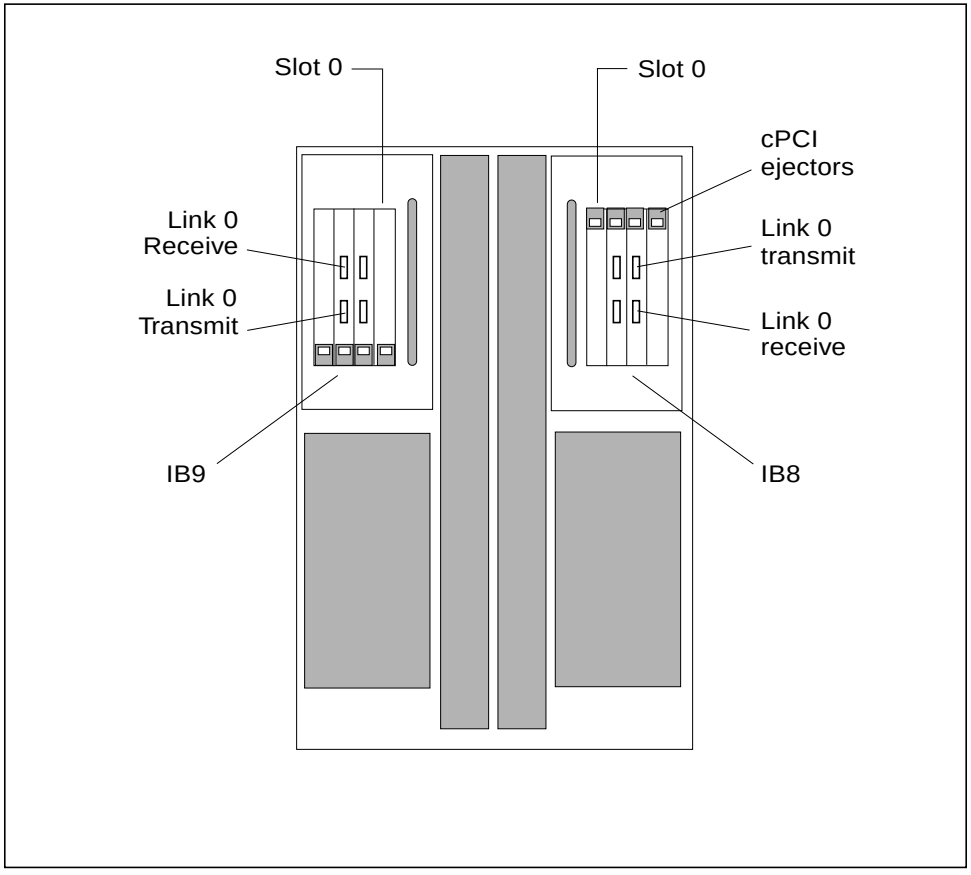

FIGURE 3-2 Sun Fire 6800 System With Sun Fire Link Assemblies

Sun Fire Link Assembly for Sun Fire 15K/12K Systems

The Sun Fire Link assembly for Sun Fire 15K/12K systems is illustrated in FIGURE 3-3.

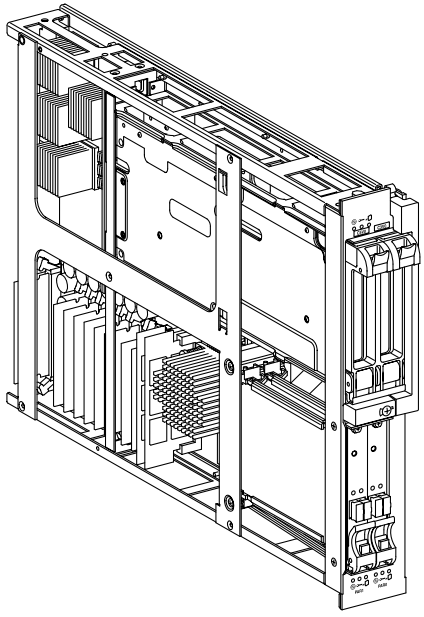

FIGURE 3-3 Sun Fire Link Assembly for Sun Fire 15K/12K System

The Sun Fire Link assembly for Sun Fire 15K/12K systems has two slots for Sun Fire Link optical modules and two slots for standard PCI adapters in hot-plug cassettes. The Sun Fire Link optical modules occupy the upper two slots.

A Sun Fire 15K/12K system can have up to four domains connected to a Sun Fire Link network, with two Sun Fire Link assemblies per domain and two optical links per assembly. Therefore, a Sun Fire 15K/12K system that has the maximum number of Sun Fire Link assemblies will have 16 Sun Fire Link connections to the Sun Fire Link network.

Sun Fire Link assemblies are installed in the Sun Fire 15K/12K I/O assembly bays. Although all 18 I/O slots are available to the Sun Fire Link assemblies, their slot assignments must adhere to a logical address pairing rule. That is, assemblies must be placed in pairs, with each pair occupying logically adjacent slots, such as 0 and 1, 2 and 3, and so forth up to 16 and 17. For example, if one Sun Fire Link assembly is installed in slot 4, a second Sun Fire Link assembly must be in slot 5.

Sun Fire Link Optical Module

Each Sun Fire Link optical module supports a single bidirectional optical link. An optical link's transceiver circuits are packaged on a cPCI form factor card, which is installed in a Sun Fire Link assembly.

Color-coded transmit and receive connectors are mounted on the module's external panel. The dark connector is used for incoming signals and the light connector for outgoing signals. The transmit and receive paths each consist of 12 simplex channels. One channel carries a clock signal and the 11 other channels carry packet framing, data, and error check codes. Each channel operates at 1.2 Gbits per second. An entire link cable operates at up to 1.65 Gbytes per second of raw bandwidth in each direction.

Sun Fire Link Optical Cables

Each optical link cable consists of a pair of 12-channel, simplex optical cables that are bundled to form a single duplex cable. The cables are available in three standard lengths: 5, 12, and 20 meters.

The cable connectors are color coded to match the optical module connectors.

Sun Fire Link Switch

The Sun Fire Link switch is an eight-port optical switch that is housed in a 19-inch, rack-mount chassis. The chassis is seven rack units high. See FIGURE 3-4.

The eight optical ports are serviced by eight Sun Fire Link optical modules of the same type used in the Sun Fire Link assemblies.

Message traffic within the switch is managed by a Sun Fire Link switch ASIC. This is a modified version of the Sun Fire Link ASIC used in the Sun Fire Link assembly. The ASIC controls eight bidirectional link paths instead of one, operating, in effect, as a crossbar switch. The Sun Fire Link switch ASIC is located on the Sun Fire Link switch assembly.

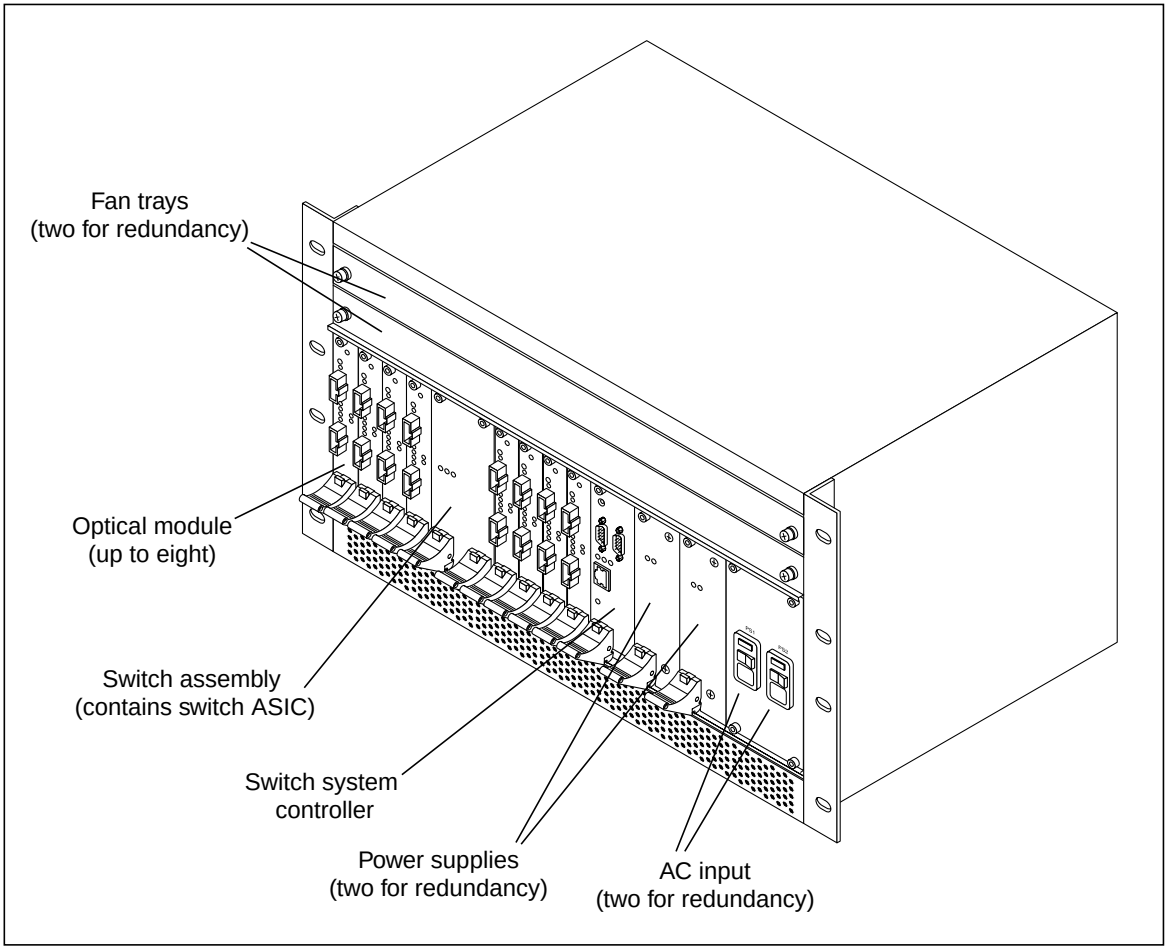

FIGURE 3-4 Sun Fire Link Switch

Local management of the switch is provided by a Sun Fire Link switch system controller, which has a resident microprocessor. The switch system controller includes one Ethernet port and two serial ports. Remote administration of the switch is exercised over the Ethernet connection.

Power to the switch is redundant, in keeping with the general high-availability design of the Sun Fire Link system. The switch has two AC power cords, each of which supports an AC/DC 150W power supply that has a cPCI form factor.

Two fan tray assemblies located above the cPCI card cage provide air cooling to the switch circuitry. Fan speed is monitored by the switch system controller.

The optical modules, switch system controller, power supplies, and fan trays are hot-swappable. That is, they can be removed and replaced while the switch has power applied.

CHAPTER 4

Overview of Sun Fire Link Software

The software components required to support the Sun Fire Link interconnect include:

- Sun Fire Link Cluster Driver Software
- Sun Fire Link Switch Software
- Sun Fire Link Administration Software

These components are described separately below.

Sun Fire Link Cluster Driver Software

The Sun Fire Link interface is managed by a device driver that implements the Remote Shared Memory (RSM) interface. RSM supports operations on remote memory as if it were local.

Sun Fire Link Switch Software

The Sun Fire Link switch system controller uses a VxWorks-based software environment. It exports the following administration services:

- Switch status is exported to the Sun Management Center, as well as to any other interested client software through SNMP.
- Switch control of certain hardware features, such as power supply levels, is exported to the Sun Management Center, or to any other interested client software. This control is also handled through SNMP.
- Fabric routing and configuration services are exported to the Sun Fire Link Manager (FM) software through private Java™ RMI interfaces.

- Any switch features that are exportable through the SNMP interface can also be accessed by an attached serial device through the switch system controller's serial interface.

Sun Fire Link Administration Software

Sun Fire Link software includes tools for administering Sun Fire Link networks. Administration of Sun Fire Link networks includes the following tasks:

- Configuring and reconfiguring Sun Fire Link partitions
- Dynamically adding nodes to and removing nodes from partitions
- Bringing up and taking down optical links
- Enforce domain topology constraints
- Monitoring a configured cluster for faults, such as link failures

These tasks are performed by the Sun Fire Link Manager module, which runs on a server external to the Sun Fire Link cluster. The Sun Fire Link Manager manages a set of hardware components (the fabric) and, in concert with a set of Sun Fire Link Manager agents running locally on the cluster nodes, carries out the administration tasks listed above.

The Sun Fire Link Manager provides a simple command-line interface that can be used to administer Sun Fire Link clusters.

It also provides a set of Sun Management Center add-on components that make available the full set of Sun Management Center controls and monitoring capabilities that are available at the Solaris and hardware levels. In particular, the Sun Management Center interface offers the following features not available in the command-line interface:

- Real-time reporting of network status
- Graphical representation of fabric components, routing information, and link status

The software components responsible for Sun Fire Link administration are discussed briefly below.

Sun Fire Link Manager

The Sun Fire Link Manager (FM) is installed on a host that is external to the Sun Fire Link cluster. It has the following responsibilities:

- Manages a set of components, which are referred to collectively as the fabric
- Determines network topologies for Sun Fire Link partitions

- Recommends changes in existing topologies
- Calculates valid configurations
- Installs configurations
- Monitors links and routes

Sun Management Center / FM Console

The Sun Management Center / FM console is the set of Sun Management Center windows and dialog screens that are specific to managing the Sun Fire Link cluster. It provides the following services:

- Provides a graphical interface to the Sun Fire Link Manager
- Collects status from the Sun Fire Link Manager and SNMP proxies
- Displays link status
- Reports on failed fabric components

The Sun Management Center / FM console ordinarily runs on the same host as the Sun Management Center server. However, it can be on a different system that is either part of the Sun Fire Link cluster or is external to it. In other words, so long as it has a network connection to the Sun Management Center server, its physical location is not important.

FM Proxies

Each standalone hardware entity that communicates with the Sun Fire Link Manager is host to an FM proxy. These are the system controllers, Solaris instances, and switches. The FM proxy running on Solaris is called the WRSM proxy.

The FM and the FM proxies have a client / server relationship where the FM is the client. FM proxies handle configuration requests from the FM. This includes the following responsibilities:

- Accepts configuration requests from the FM
- Forwards FM requests to the local Sun Fire Link device driver
- Stores the current configuration for use on reboot

Sun Management Center Server

The Sun Management Center server supports monitoring and control activities being performed by the Sun Management Center / FM console, configuration, and monitoring modules. In particular, it forwards user input and system events to the appropriate modules.

The Sun Management Center server runs on a Solaris host that *should* be external to the Sun Fire Link cluster for performance reasons, but need not be.

Sun Fire Link Device Drivers

Sun Fire Link device drivers manage the local Sun Fire Link ASICs and link interfaces on the system controllers, switches, and Solaris instances. These device drivers have the following responsibilities:

- Communicate with other Sun Fire Link device drivers to identify link endpoints
- Install route maps
- Program striping levels
- Monitor links and errors
- Generate link and routing events

SNMP Agents

SNMP agents are responsible for collecting network status for the Sun Management Center console. They have the following responsibilities:

- Polling for status
- Notifying Sun Management Center of events with SNMP traps
- Returning information requested by Sun Management Center

There are two types of SNMP agent:

- Sun Management Center agents – These agents are installed on Solaris and are based on Sun Management Center technology
- Java Dynamic Management™ Kit agents – These agents are installed on embedded systems (node system controllers and switch system controllers) and are based on Java.

Sun Management Center Agent/System Controller Proxy, Switch Proxy

There is a Sun Management Center proxy for each non-Sun Management Center agent. These proxies provide interfaces between Sun Management Center and the SNMP agents for the system controller and switch system controller.

Index

A

alternate path, *See* failover
ASIC, Sun Fire Link, 13
assembly, *See* Sun Fire Link assembly.

C

cluster, definition, 2
command-line interface, 20
configurations
  direct-connect, 2, 6
  four-switch, 8
  overview, 5 to 10
  Sun Fire 6800 and 15K/12K systems, 5
  Sun Fire Link switch, 2
  switched, 7
  two-node, 7
  two-switch, 8

D

device drivers, 19, 22
direct-connect configuration, 2
domains
  number in Sun Fire 15K/12K system, 6
  number in Sun Fire 6800 system, 5
driver software, *See* device drivers.

F

fabric, 4
failover, 3
  striping levels, 9
four-switch configurations, 8
four-way striping, 9

G

graphical interface, 21

H

hardware
  I/O slots in a Sun Fire 15k/12K system, 15
  I/O slots in a Sun Fire 6800 system, 13
  Sun Fire Link ASIC, 13
  Sun Fire Link assembly, 2
    for Sun Fire 15K/12K systems, 14
    for Sun Fire 6800 systems, 11
  Sun Fire Link optical module, 2, 16
  Sun Fire Link switch, 3, 16

I

I/O slots
  Sun Fire 15K/12K systems, 15
  Sun Fire 6800 systems, 13
IB8 and IB9
  illustration, 14

slots in a Sun Fire 6800 system, 13

L

level-four striping, 9
level-two striping, 9
link failures, 3

M

Management Station, 4

N

number of domains, 5

O

optical module, *See* Sun Fire Link optical module.

P

partition
  definition, 4, 9
proxies, 22

R

rack units, switch, 16
Remote Shared Memory (RSM), 2

S

SNMP agents, 22
software
  administration, 20
  command-line interface, 20
  device drivers, 19, 22
  FM proxy, 21
  graphical interface, 21
  overview, 19
  SNMP agents, 22
  Sun Fire Link Manager, 4, 20
  Sun Fire Link switch, 19
  Sun HPC ClusterTools applications, 9
  Sun Management Center, 19
    proxy, 22
    server, 21
    Sun Management Center/FM console, 21
striping
  hardware, 3
  levels of striping, 9
  limitation in three-node direct-connect configurations, 10
Sun Fire 15K/12K system
  I/O slots, 15
  number of domains, 6
Sun Fire 6800 system, 13
  I/O slots, 13
  number of domains, 5
Sun Fire Link ASIC, 13
Sun Fire Link assembly for Sun Fire 15K/12K systems, 14
  slots available, 15
Sun Fire Link assembly for Sun Fire 6800 systems, 11
  block diagram, 12
  illustration, 12
  slots available, 13
Sun Fire Link cluster
  definition, 2
Sun Fire Link Manager
  command-line interface, 20
  description, 20
  Ethernet connection, 4
  Managment Station, 4
  proxy, 21
Sun Fire Link optical module, 2, 16
Sun Fire Link switch
  configurations, 2, 8
  hardware, 16
  height, 16
  rack units, 16
  software, 19
  switch ASIC, 16
Sun HPC ClusterTools applications, 9
Sun Management Center, 19
  proxy, 22

server, 21
Sun Management Center/FM console, 21
switch, *See* Sun Fire Link switch.

T

three-node configuration, 7
striping limitation in direct-connect mode, 10
two-node configuration, 7
two-switch configurations, 8
two-way striping, 9