STORAGETEK®

FLEXLINE™

ONLINE STORAGE SYSTEMS

INSTALLATION INSTRUCTIONS FOR IRIX
Open Systems Host Attach

PRODUCT TYPE
SOLUTIONS

FlexLine™ 200/300 Series

Installation Instructions for IRIX open systems disk host attach

Information Control

The information in this document, including any associated software program, may not be reproduced, disclosed or distributed in any manner without the written consent of Storage Technology Corporation.

Should this publication be found, please return it to StorageTek, One StorageTek Drive, Louisville, CO 80028-5214, USA. Postage is guaranteed.

Information in this file is subject to change. In the event of changes, the information will be revised. Comments concerning its contents should be directed to:

200/300 Series Program Manager
Storage Technology Corporation
One StorageTek Drive
Louisville, CO 80028-2129

or send email to: **glsfs@stortek.com**

Export Destination Control Statement

These commodities, technology or software were exported from the United States in accordance with the Export Administration Regulations. Diversion contrary to U.S. law is prohibited.

Proprietary Information Statement

This document and its contents are proprietary to Storage Technology Corporation and may be used only under the terms of the product license or nondisclosure agreement. The information in this document, including any associated software program, may not be reproduced, disclosed or distributed in any manner without the written consent of Storage Technology Corporation.

Copyright Statement

Summary of Changes

The following is the history and summary of changes for this publication.

Date	Edition	Description
October 2003	First	Initial Release
October 2004	Second	Updated to reflect new FlexLine™ terminology

This page intentionally left blank.

IRIX - Installation Instructions

1

This manual is organized to help plan your IRIX configuration, provide order information, and help to install and attach your FlexLine™ 200/300 Series storage systems to the host server running IRIX.

■ Introduction

The following should be used in conjunction with the Siebel configurator.

Plan your site configuration

Prior to placing an order for StorageTek storage equipment, one should make sure of the system needs. First measure the locations of each current and future host, storage system, and other network components. Fill in this form with the requested information.

✔	Parameter	Outcome
	IRIX 6.5.18 operating level or better and SGI Patch 4417	
	Distance between host and storage system	
	Distance between other devices	
	Number of HBAs needed (two are recommended)	
	Number of storage systems being installed	
	Number of FC cables needed	

Refer also to the Technical Qualification Worksheet. This system assurance tool should be used to ensure site readiness.

Select Host Adapter (s)

Pick the type and quantity of the QLogic fibre channel PCI Host Bus Adapters (HBAs) needed for your host systems.

Quantity	Description	Rating
	QLogic, 2310F/66 - Multi-mode optic, SFP Transceiver interface	2.0 Gb/sec

Select the Fibre Channel Cables

Select the length of fiber optic cable and the quantity for connection to the host systems. The following is an example of optical cables that may be ordered through StorageTek to attach a QLA2310F HBA to the Control Module (aka Controller). Write in the quantities needed.

Quantity	2 Gb FC Optical Cable Description	Part #
	10 meters (32.8 feet) LC to LC Duplex	10800221
	50 meters (164 feet) LC to LC Duplex	10800222
	100 meters (328 feet) LC to LC Duplex	10800223
Note: Other lengths are available. Consult the Siebel Configurator.		

There are also LC to SC cables that would allow one to use a 2 Gb HBA and cables with existing switches. The following cables are available in this event.

Quantity	FC Optical Cable Description	Part #
	10 meters (32.8 feet) LC to SC	10800231
	50 meters (164 feet) LC to SC	10800232
	100 meters (328 feet) LC to SC	10800233

If installing a switch, you may be installing a set of the following.

Quantity	1 Gb FC Optical Cable Description	Part #
	10 meters (32.8 feet) SC to SC Duplex	10800134
	50 meters (164 feet) SC to SC Duplex	10800137
	100 meters (328 feet) SC to SC Duplex	10800138
Note: Other lengths are available. Consult the Seibel Configurator.		

Check your quantities against your earlier distance measurements to ensure your order includes all cable attachments. One end of the fibre channel cable mates with the HBA, and the other end, typically, mates with the control module (controller) host interface.

Note: Unless there is an existing management station, cable connections to the management station should be via 10BaseT Ethernet.

LC to LC Duplex Fibre Channel short wave optical cables are used with Small Form-factor Packaged (SFP) Transceivers to attach to the host minihubs.

Place your Order

Make sure that you have the following.

✔	Description
	Host Server running IRIX 6.5.18 or better
	Qlogic 2310F HBA(s)
	Fibre Channel cables to configure your storage system

When you make a new order, a CD-ROM along with a CD insert, commonly called a Quick Start Guide will be in your host attach kit. The Quick Start Guide will walk you through your installation, referring you to the Preconfiguration and Maintenance Set-up Instructions, P/N 95966, and then to any special instructions written to support your particular operating system installation.

CD-ROM file structure

During your installation you will be asked to install files from the CD supplied in your host-attach kit. The CD is structured (in most cases) with three main folders (Docs, Software, and Drivers). Inside the Docs folder are subfolders with pertinent documentation. The following table shows the subfolder names.

Sales and Preinstall	Diagnostic and Service	Engineering Papers
Hardware Install	Cabinet Docs	Miscellaneous
Software Install	Release Notes	

Inside the Software folder you will find four compressed files containing: SMclient, SMagent, SMruntime, and SMutil applications.

The HBA drivers are contained within IRIX, and no information regarding installation nor configuration is contained in the Drivers folder.

Note: If you need information about installing HBAs, drivers, or cables; refer to the Client Server Disk installation supplement, P/N 96015, or to your host installation manuals.

Starting with version 8.3, a special SMclient (Field Tool) CD is available and is required to fulfill your preconfiguration requirements. It has the software tools and applications that may be necessary to upgrade your storage system.

> **Note:** Refer also to the PDF documents and configuration and support files to assist with your installation.

■ Installation

Once you have ordered your hardware and software, read the following information on how to attach the 200/300 Series storage system to a server running IRIX 6.5.18 or better, and including SGI Patch 4417.

> **Note:** The SANtricity Storage Manager version 9.1 Installation Guide for IRIX (p/n ES14722) is available to assist with your software installation. Read this document ahead of time. It is available on the CD, on the CRC (Current Products >> Software) and on a backup site at: http://gandalf.stortek.com/ctp/flexline/

Before You Begin

The materials needed for the installation are listed below. Check to see that you have each of these items. The following pages provide more information about these materials (see headings) with specific steps and details on what to do.

In most cases, actual commands have been provided to help you with your installation, or help you configure or modify settings within the host or on the 200/300 Series storage system.

Before you begin, make sure you have these items:

1. A server running IRIX 6.5.18 or greater, and SGI Patch 4417

2. One or more QLogic HBA(s) - model 2310F

3. SFP Transceivers for the storage system

4. Fibre Channel cable(s) that will connect the host to the storage system

5. A storage system that is fully configured.

> **Note:** If you don't have your storage system configured, you can attach your laptop running SMclient to it. This will enable you to assign drives, LUNs, and volumes.

■ Storage System and Storage Manager - Overview

The FlexLine™ 200/300 Series storage systems provide access to high-end storage using Fibre Channel and RAID technology. A smorgasbord of hosts and applications are available to help ensure the very best in data access and acquisition - no matter how complex your network configuration.

This document introduces you to the latest in StorageTek's ongoing enhancements in Client Server Disk technology, as well as the enhancements that the storage system can provide to the IRIX operating system.

In addition to this manual, both hardware and software information is available on the engineering web sites at:

http://gandalf.stortek.com/ctp/index.html

http://gandalf.stortek.com/ctp/manager/index.html

http://gandalf.stortek.com/ctp/installation/index.html

http://gandalf.stortek.com/ctp/cabinet/index.html

NVSRAM, controller/ESM firmware, and various other files may be downloaded from the CRC by going to **Current Products >> Disk** and selecting the storage system you want.

The following information is covered in this document to help you attach the storage system to the SGI host.

This document also describes the prerequisites and steps necessary to attach a 200/300 Series storage system to an SGI running IRIX. All testing was done using an Origin 2000 with the required hardware and software described in the following sections.

■ Installation Preparation

The following sections provide the things you should know to prepare and install a storage system. Please read this information to minimize installation problems.

Hardware

IRIX currently supports the SGI XIO fibre HBA (xio) and the Qlogic 2300 HBA (pci). Some of the older XIO boards (e.g. SG-XT-FC-2P) will NOT work with the new storage systems. It (or its driver) violates the FC/AL protocol by ignoring the receive buffer size in the PLOGI exchange and defaulting to 2k.

The newer XIO boards (e.g. SG-XT-FC-1POPT), however, have been reported to work in the field.

The Qlogic HBA is the tested and supported solution for the 200/300 Series product solutions. It requires PCI slots. In the Origin 2000, we needed the XIO to PCI card cage assembly (see illustrations below) to provide the PCI slots. The Qlogic cards used are the same as what is distributed in the Windows host attach kits.
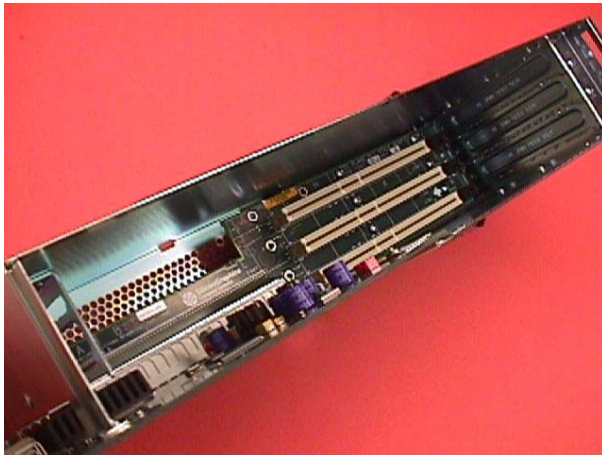
Install your Qlogic QLA2310F HBAs, plug the card cage assembly into the back of the SGI host running IRIX.

Note: Do NOT run your fibre channel cables to the storage system at this time. Preconfiguration steps must be done prior to attaching the controllers.

Proceed to the next section.

Software

IRIX has built-in drivers for the Qlogic QLA2310F and the XIO to PCI card cage. The minimum versions required for IRIX are:

    IRIX 6.5.18

Refer to the Interop Tool for connectivity information. It is available on the Sales Community. For additional information refer the SANtricity Storage Manager version 9.1 for IRIX Installation Guide (p/n ES14722).

■ Configuration

The user needs to determine whether or not they are going to use failover. The decision (with respect to the customer's use of failover) will come into play at the time that the volumes are mapped to the host ports.

Failover Decision

It should be noted that the IRIX failover requires the LUNs to be XLV volumes — so they need to be set up that way in the beginning.

NVSRAM

The NVSRAM load for heterogeneous includes failover and non-failover modules. Contrary to earlier code releases, AVT is enabled. The primary distinction between failover and non-failover options comes from the path representation to the host. In a non-failover installation, the secondary path is masked.

Preconfiguration

The standard method of preconfiguring the storage system should be followed prior to attaching to the host. Refer to the Preconfiguration and Maintenance Setup Instructions manual p/n 95966.

Using the SMclient field tool 8.4 or 9.1 (or better) on a PC attached to the ethernet ports, download the NVSRAM and firmware (if needed) to the controllers.
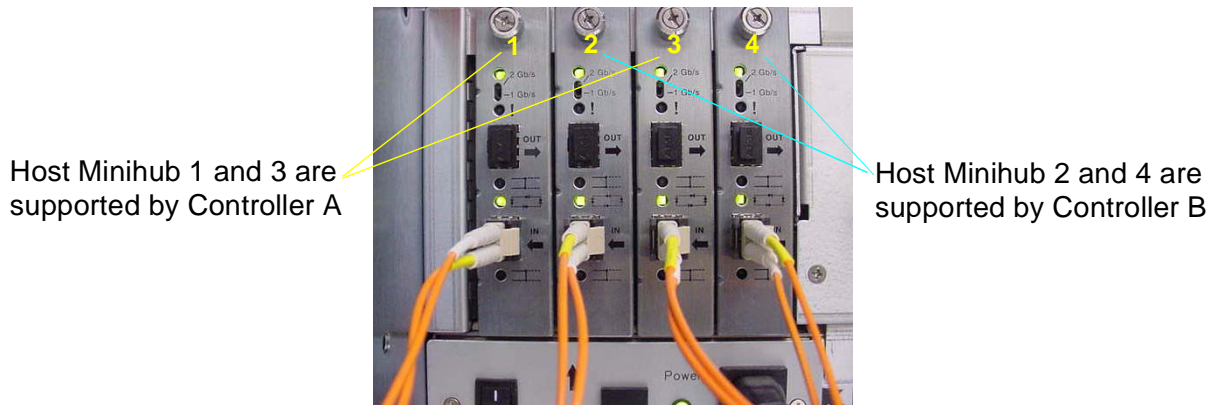
Build the RAID groups and LUNs.

Note: Previous products had cache mirroring disabled by default when LUNs are created on an SGI platform to further increase write performance in graphics intensive operations. The default is to have cache mirroring enabled to protect write cache data in the event of a controller failure. It is up to the customer to decide whether to trade off this protection for better write performance.

Attach to the host and reboot.

Note: Make sure that your host minhub speed is set to correspond with your HBA and switch speed. See photograph below.
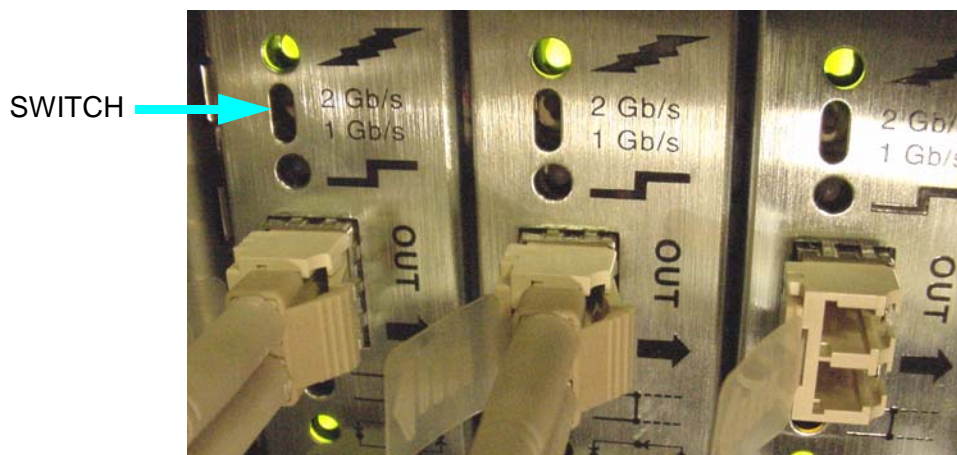
Run your fibre channel cables to separate controllers on the storage system. You may run FC cables through a switch or directly to the controllers. See photograph below.

Host Minihub 1 and 3 are supported by Controller A

Host Minihub 2 and 4 are supported by Controller B

What you should see directly connected

The following are examples of the types of things you should see when the 200/300 Series storage system is attached and operating correctly. Two cases are presented here for clarification.

Note: There is a substantial distinction in the device respresentation when comparing a switched and a non-switched environment. Both cases are built on the intent to have failover.

Non-switched attach

The storage system used had 20 drives configured into four RAID 5 groups. LUNs 0 and 2 were assigned to controller A; and LUNs 1 and 3 were assigned to controller B. Controller A is attached to QL2200A controller 11; and controller B is attached to QL2200A controller 10.

hinv -v command

Two 195 MHZ IP27 Processors

CPU: MIPS R10000 Processor Chip Revision: 2.6

FPU: MIPS R10010 Floating Point Chip Revision: 0.0

CPU 0 at Module 1/Slot 1/Slice A: 195 Mhz MIPS R10000 Processor Chip (enabled)

Processor revision: 2.6. Secondary cache: Size 4 MB Speed 130 Mhz

CPU 1 at Module 1/Slot 1/Slice B: 195 Mhz MIPS R10000 Processor Chip (enabled)

Processor revision: 2.6. Secondary cache: Size 4 MB Speed 130 Mhz

Main memory size: 256 Mbytes

Instruction cache size: 32 Kbytes

Data cache size: 32 Kbytes

Secondary unified instruction/data cache size: 4 Mbytes

Memory at Module 1/Slot 1: 256 MB (enabled)

Bank 0 contains 128 MB (Standard) DIMMS (enabled)

Bank 1 contains 128 MB (Standard) DIMMS (enabled)

Integral SCSI controller 0: Version QL1040B (rev. 2), single ended

Disk drive: unit 1 on SCSI controller 0 (unit 1)

CDROM: unit 6 on SCSI controller 0

Integral SCSI controller 1: Version QL1040B (rev. 2), single ended

Integral SCSI controller 10: Version Fibre Channel QL2200A

Disk drive: unit 5 on SCSI controller 10 (unit 5)

Disk drive: unit 5, LUN 1 on SCSI controller 10 (unit 5)

Disk drive: unit 5, LUN 2 on SCSI controller 10 (unit 5)

Disk drive: unit 5, LUN 3 on SCSI controller 10 (unit 5)

Integral SCSI controller 11: Version Fibre Channel QL2200A

Disk drive: unit 2 on SCSI controller 11 (unit 2)

Disk drive: unit 2, LUN 1 on SCSI controller 11 (unit 2)

Disk drive: unit 2, LUN 2 on SCSI controller 11 (unit 2)

Disk drive: unit 2, LUN 3 on SCSI controller 11 (unit 2)

Integral SCSI controller 4: Version QL1040B (rev. 2), differential

Integral SCSI controller 5: Version QL1040B (rev. 2), differential

Integral SCSI controller 6: Version QL1040B (rev. 2), differential

Integral SCSI controller 7: Version QL1040B (rev. 2), differential

Integral SCSI controller 8: Version Fibre Channel AIC-1160, revision 1

Integral SCSI controller 9: Version Fibre Channel AIC-1160, revision 1

IOC3 serial port: tty1

IOC3 serial port: tty2

IOC3 serial port: tty3

IOC3 serial port: tty4

IOC3 serial port: tty5

IOC3 serial port: tty6

IOC3 serial port: tty7

IOC3 serial port: tty8

Integral Fast Ethernet: ef0, version 1, module 1, slot io1, pci 2

Fast Ethernet: ef1, version 1, module 1, slot io4, pci 0

Fast Ethernet: ef2, version 1, module 1, slot io4, pci 1

Fast Ethernet: ef3, version 1, module 1, slot io4, pci 2

Fast Ethernet: ef4, version 1, module 1, slot io4, pci 3

Origin BASEIO board, module 1 slot 1: Revision 3

    PCI Adapter ID (vendor 4265, device 3) pci slot 2

    PCI Adapter ID (vendor 4215, device 4128) pci slot 0

    PCI Adapter ID (vendor 4215, device 4128) pci slot 1

Origin MENET board, module 1 slot 4: Revision 3

    PCI Adapter ID (vendor 4265, device 3) pci slot 0

    PCI Adapter ID (vendor 4265, device 3) pci slot 1

    PCI Adapter ID (vendor 4265, device 3) pci slot 2

    PCI Adapter ID (vendor 4265, device 3) pci slot 3

Origin PCI XIO board, module 1 slot 2: Revision 4

    PCI Adapter ID (vendor 4215, device 8704) pci slot 0

    PCI Adapter ID (vendor 4215, device 8704) pci slot 1

Origin MSCSI board, module 1 slot 5: Revision 3

    PCI Adapter ID (vendor 4215, device 4128) pci slot 0

    PCI Adapter ID (vendor 4215, device 4128) pci slot 1

    PCI Adapter ID (vendor 4215, device 4128) pci slot 2

    PCI Adapter ID (vendor 4215, device 4128) pci slot 3

Origin FIBRE CHANNEL board, module 1 slot 3: Revision 3

    PCI Adapter ID (vendor 36868, device 4448) pci slot 0

PCI Adapter ID (vendor 36868, device 4448) pci slot 1

IOC3 external interrupts: 1

HUB in Module 1/Slot 1: Revision 5 Speed 97.50 Mhz (enabled)

IP27prom in Module 1/Slot n1: Revision 6.60

IO6prom on Global Master Baseio in Module 1/Slot io1: Revision 6.60

Note the two QL2200A cards with all 4 LUNs reported under each. You need to note the controller, unit, and LUN numbers; and also have available which LUNs are assigned to which controller for the failover configuration. In this example, the following paths to the 4 LUNs exist.

Table 1. LUN Paths

LUN	Primary controller/unit	Alternate controller/unit
0	11/2	10/5
1	10/5	11/2
2	11/2	10/5
3	10/5	11/2

Also note that LUN 0 is listed, but not stated explicitly like the other LUNs, i.e.:

Disk drive: unit 5 on SCSI controller 10 (unit 5)
Disk drive: unit 5, LUN 1 on SCSI controller 10 (unit 5)

These devices will show up prior to labeling them in /dev/scsi as shown.

Note: The sc0 LUNs are the system disk.

sc0d1l0   sc10d5l0   sc10d5l2   sc11d2l0   sc11d2l2

sc0d6l0   sc10d5l1   sc10d5l3   sc11d2l1   sc11d2l3

If failover is not being used, there will only be one entry on the path that is assigned.

Label and set parameters with fx –x

The customer will need to partition and label the disks using the fx command. Configure only the primary path. There are 4 physical LUNs in this example and 8 paths. In the above example, you would run fx –x and configure the following 4:

Prompt	Response	LUN 0	LUN 1	LUN 2	LUN 3
Fx: "device name" =(dksc)	Enter				
Fx: ctlr# = (0)		10	11	10	11

Prompt	Response	LUN 0	LUN 1	LUN 2	LUN 3
Fx: drive# = (1)		5	2	5	2
Fx: lun# = (0)		0	1	2	3

The only non standard setting the storage system needs is to set the command queuing on. This is accomplished under the path [l]abel->[se]t->[para]meters. Change the Enable CTQ = (disabled) to enabled and CTQ depth to the lesser value of 16 or 256/# of LUNs. Keep future expansion in mind. A value of 8 will allow 32 LUNs to be configured.

After configuration, the /dev/dsk directory should contain the slices of these devices. These were configured with the default of 2 slices (s0 and s1):

dks0d1s0   dks10d5l1s0 dks10d5l3s0 dks11d2l2s0 dks11d2s0   root

dks0d1s1   dks10d5l1s1 dks10d5l3s1 dks11d2l2s1 dks11d2s1   swap

Note how LUN 0 (dks11d2s0) does not have the l0 like other LUNs do.

Configuring failover using XLV volume manager

Examine the following man pages for an understanding of the failover methodology implemented in IRIX:

Man failover

Man xlv_mgr

Man xlv_make

Man scsifo

File /etc/failover.conf

The failover mechanism requires the use of the XLV volume manager. The volumes should be set up using the `xlv_make` command and supply the primary path name to the ve parameter. The configured volumes will show up in `/dev/xlv`.

The alternate paths need to be set up in the `/etc/failover.conf` file. It is useful to turn on the #verbose mode during setup. It may also be necessary to set the #disable_target_lun_check mode. By default, IRIX will not allow one to assign a path as an alternate unless the target and LUN numbers match. The 200/300 Series controllers will most likely be on different target addresses, so the directive needs set.

Here is the end of the file for the example:

#verbose

#disable_target_lun_check

#Name Pri path   sec path

Lun00 sc11d2l0 sc10d5l0

Lun01 sc10d5l1 sc11d2l1

Lun02 sc11d2l2 sc10d5l2

Lun03 sc10d5l3 sc11d2l3

The name field is arbitrary and the paths come from the /dev/scsi. As noted in the man pages and the failover.conf file, if it doesn't exist in /dev/scsi, you can't use it.

Here is the text portion of `/etc/failover.conf:`

#ident $Revision: 1.1 $

\#

\# This is the configuration file for table driven failover support.

\# The entries within this file consist of a single line per

\# failover grouping. These lines describe up to four paths to

\# each device. Obviously, these paths should reference the

\# same device! Some sanity checking is performed, but the

\# software is not foolproof in this area. (If it was foolproof,

\# the configuration file would be unnecessary!) It is recommended

\# that all paths to a device be listed.

\#

\# The format of each line consists of a group name followed by

\# up to four device names from the "/dev/scsi" directory. If

\# it's not in "/dev/scsi", it cannot be configured. But, this

\# implementation assumes that if a device is not present, its

\# path has failed. Consequently, no error messages will be

\# generated for missing paths. This means that you may have a

\# a failover group that consists of a single primary path with

\# no alternates.

\#

\# Additionally, there is a configuration directive available that

\# will cause the program to emit debug information. Placing

\# "#verbose" at the start of a line, without the quotes, will

```
#        enable the debug output.  The debug output is displayed for
#        all subsequent configuration directives.  Once enabled, the
#        debug output cannot be disabled.
#
#        Lines that begin with a "#" are considered comment lines.
#        A "#" anywhere within a line signals the beginning of a comment.
#        White space must separate the last parameter of a line and the #.
#        Blank lines are also considered a comment.
#
#        --> Lines that begin with 'sc' will be skipped as it's likely a
#        --> configuration error.
#
#        Sample configuration.  (Remember to omit the "#" the start of
#        the line.)
#
#
# #      Name    Pri path | alt path | alt path | alt path
#
#        GroupA  sc6d50l0   sc7d50l0   sc8d50l0   sc9d50l0
#        GroupB  sc6d51l0   sc7d51l0   sc8d51l0   sc9d51l0 # Some comment
#        GroupC  sc6d52l0   sc7d52l0   sc8d52l0   sc9d52l0
#
#        One word of caution.  The switch to an alternate path is
#        choreographed by XLV.  XLV is a requirement for failover to
#        function.
#
#        Notes:
#
#        It is not possible to change the primary device of a group after
#        the initial configuration by reordering the entries on the line
```

Fourth Edition

and running /etc/init.d/failover start. To change the primary,

scsifo may be used to switch to the next available path.

#

It is not possible to remove the primary of a group by deleting

the entry from the group and running /etc/init.d/failover start.

A reboot is necessary.

#

To remove the "DOWN" displayed by hinv for the failed path,

fix the path and reprobe the bus using scsiha -p #.

After this file has been set up and saved, run "`/etc/init.d/failover start`" or reboot. The console comes up with text like this when in verbose mode:

Configuring Failover.

Processing:

Processing: #disable_target_lun_check

Processing: #Name Pri path   sec path

Processing:

Processing: Lun00 sc11d2l0 sc10d5l0

/sbin/foconfig: candidate: sc11d2l0 -> /hw/module/1/slot/io2/pci_xio/pci/1/scsi_ctlr/0/target/2/lun/0/scsi

/sbin/foconfig: candidate: sc10d5l0 -> /hw/module/1/slot/io2/pci_xio/pci/0/scsi_ctlr/0/target/5/lun/0/scsi

Processing: Lun01 sc10d5l1 sc11d2l1

/sbin/foconfig: candidate: sc10d5l1 -> /hw/module/1/slot/io2/pci_xio/pci/0/scsi_ctlr/0/target/5/lun/1/scsi

/sbin/foconfig: candidate: sc11d2l1 -> /hw/module/1/slot/io2/pci_xio/pci/1/scsi_ctlr/0/target/2/lun/1/scsi

Processing: Lun02 sc11d2l2 sc10d5l2

/sbin/foconfig: candidate: sc11d2l2 -> /hw/module/1/slot/io2/pci_xio/pci/1/scsi_ctlr/0/target/2/lun/2/scsi

/sbin/foconfig: candidate: sc10d5l2 -> /hw/module/1/slot/io2/pci_xio/pci/0/scsi_ctlr/0/target/5/lun/2/scsi

Processing: Lun03 sc10d5l3 sc11d2l3

/sbin/foconfig: candidate: sc10d5l3 -> /hw/module/1/slot/io2/pci_xio/pci/0/scsi_ctlr/0/target/5/lun/3/scsi

/sbin/foconfig: candidate: sc11d2l3 -> /hw/module/1/slot/io2/pci_xio/pci/1/scsi_ctlr/0/target/2/lun/3/scsi

Processing:

Processing:

Processing:

Calling xlv_assemble.

If all is successful, the `hinv -v` and the `scsifo -d` command will reflect the primary and alternate paths.

The hinv section looks like:

two 195 MHZ IP27 Processors

CPU: MIPS R10000 Processor Chip Revision: 2.6

FPU: MIPS R10010 Floating Point Chip Revision: 0.0

CPU 0 at Module 1/Slot 1/Slice A: 195 Mhz MIPS R10000 Processor Chip (enabled)

    Processor revision: 2.6. Secondary cache: Size 4 MB Speed 130 Mhz

CPU 1 at Module 1/Slot 1/Slice B: 195 Mhz MIPS R10000 Processor Chip (enabled)

    Processor revision: 2.6. Secondary cache: Size 4 MB Speed 130 Mhz

Main memory size: 256 Mbytes

Instruction cache size: 32 Kbytes

Data cache size: 32 Kbytes

Secondary unified instruction/data cache size: 4 Mbytes

Memory at Module 1/Slot 1: 256 MB (enabled)

    Bank 0 contains 128 MB (Standard) DIMMS (enabled)

    Bank 1 contains 128 MB (Standard) DIMMS (enabled)

Integral SCSI controller 0: Version QL1040B (rev. 2), single ended

    Disk drive: unit 1 on SCSI controller 0 (unit 1)

    CDROM: unit 6 on SCSI controller 0

Integral SCSI controller 1: Version QL1040B (rev. 2), single ended

Integral SCSI controller 8: Version Fibre Channel AIC-1160, revision 1

Integral SCSI controller 10: Version Fibre Channel QL2200A

Disk drive: unit 5 on SCSI controller 10 (unit 5) (alternate path)

Disk drive: unit 5, LUN 1 on SCSI controller 10 (unit 5) (primary path)

Disk drive: unit 5, LUN 2 on SCSI controller 10 (unit 5) (alternate path)

Disk drive: unit 5, LUN 3 on SCSI controller 10 (unit 5) (primary path)

Integral SCSI controller 11: Version Fibre Channel QL2200A

Disk drive: unit 2 on SCSI controller 11 (unit 2) (primary path)

Disk drive: unit 2, LUN 1 on SCSI controller 11 (unit 2) (alternate path)

Disk drive: unit 2, LUN 2 on SCSI controller 11 (unit 2) (primary path)

Disk drive: unit 2, LUN 3 on SCSI controller 11 (unit 2) (alternate path)

And the `scsifo -d` looks like:

Lun00:

[P] sc11d2l0 (587)

[ ] sc10d5l0 (790)

Lun01:

[P] sc10d5l1 (587)

[ ] sc11d2l1 (790)

Lun02:

[P] sc11d2l2 (587)

[ ] sc10d5l2 (790)

Lun03:

[P] sc10d5l3 (587)

[ ] sc11d2l3 (790)

The [P] indicates the current path.

Testing failover

The following describe testing the path failover and the steps to recovering after the path is restored:

A path failure was caused by removing the fibre cable from controller A while doing I/O to all 4 LUNs in the example. The Qlogic card waits 240 seconds for a timeout to occur which then gets reported to XLV and the failover is

performed. This is quite a long time. The variable that controls this is in `/var/sysgen/master.d/qlfc` and is:

- qlfc_target_missing_timeout:

Note: Time is the seconds that the driver waits before assuming the worst about a target and begins to reject requests. Default is 240 seconds, i.e., 3 minutes. This is a reasonable value should you decide to reboot a switch with i/o going to its targets, which BTW, is something that you would never do. A good minimum for a failover environment might be 15 seconds. The lower bound, as enforced by the driver, is 2 * qlfc_watchdog_time, or 10 seconds, whichever is greater.

int   qlfc_target_missing_timeout = 240;

We recommend setting it to 20 and recompiling the kernel (see man autoconfig).

The hinv –v output shows the failed paths:

Two 195 MHZ IP27 Processors

CPU: MIPS R10000 Processor Chip Revision: 2.6

FPU: MIPS R10010 Floating Point Chip Revision: 0.0

CPU 0 at Module 1/Slot 1/Slice A: 195 Mhz MIPS R10000 Processor Chip (enabled)

  Processor revision: 2.6. Secondary cache: Size 4 MB Speed 130 Mhz

CPU 1 at Module 1/Slot 1/Slice B: 195 Mhz MIPS R10000 Processor Chip (enabled)

  Processor revision: 2.6. Secondary cache: Size 4 MB Speed 130 Mhz

Main memory size: 256 Mbytes

Instruction cache size: 32 Kbytes

Data cache size: 32 Kbytes

Secondary unified instruction/data cache size: 4 Mbytes

Memory at Module 1/Slot 1: 256 MB (enabled)

  Bank 0 contains 128 MB (Standard) DIMMS (enabled)

  Bank 1 contains 128 MB (Standard) DIMMS (enabled)

Integral SCSI controller 0: Version QL1040B (rev. 2), single ended

  Disk drive: unit 1 on SCSI controller 0 (unit 1)

  CDROM: unit 6 on SCSI controller 0

Integral SCSI controller 1: Version QL1040B (rev. 2), single ended

Integral SCSI controller 10: Version Fibre Channel QL2200A

 Disk drive: unit 5 on SCSI controller 10 (unit 5) (primary path)

 Disk drive: unit 5, LUN 1 on SCSI controller 10 (unit 5) (primary path)

 Disk drive: unit 5, LUN 2 on SCSI controller 10 (unit 5) (primary path)

 Disk drive: unit 5, LUN 3 on SCSI controller 10 (unit 5) (primary path)

Integral SCSI controller 11: Version Fibre Channel QL2200A

 Disk drive: unit 2 on SCSI controller 11 (unit 2) (alternate path) DOWN

 Disk drive: unit 2, LUN 1 on SCSI controller 11 (unit 2) (alternate path)

 Disk drive: unit 2, LUN 2 on SCSI controller 11 (unit 2) (alter'te path) DOWN

 Disk drive: unit 2, LUN 3 on SCSI controller 11 (unit 2) (alternate path)

Note the "DOWN" indication on 2 of the paths. The alternate paths and any primary paths on the "broken" path that have not had any failing I/Os issued will not failover until they do. If you pull a cable with no I/O going on to a LUN, it will not fail over until an I/O fails.

The `scsifo -d` output:

 Lun00:

 [ ] sc11d2l0 (587)

 [P] sc10d5l0 (790)

 Lun01:

 [P] sc10d5l1 (587)

 [ ] sc11d2l1 (790)

 Lun02:

 [ ] sc11d2l2 (587)

 [P] sc10d5l2 (790)

 Lun03:

 [P] sc10d5l3 (587)

 [ ] sc11d2l3 (790)

The [P] indicates the current path. Lun00 and Lun02 are now on the secondary path.

Recovering from a failover condition

To recover from a failover condition and again use the preferred paths, follow the procedure below:

1. Repair the failing path condition, i.e. cable, GBIC/SFP Transceiver, controller, or HBA

2. Run scsiha –p nn where nn is the controller path indicated as DOWN (11 in example)

3. Run hinv –v and verify DOWN condition is corrected.

4. Run scsifo –d and note any devives not on the primary path (first in list).

5. Run scsifo –s <path> on devices with [P] in secondary path with <path> being the secondaries indicated in step 4. For the example:

   Scsifo –s sc10d5l0

   Scsifo –s sc10d5l2

6. Run scsifo –d – All should be back on the primary path (top).

   Lun00:

       [P] sc11d2l0 (587)

       [ ] sc10d5l0 (790)

   Lun01:

       [P] sc10d5l1 (587)

       [ ] sc11d2l1 (790)

   Lun02:

       [P] sc11d2l2 (587)

       [ ] sc10d5l2 (790)

   Lun03:

       [P] sc10d5l3 (587)

       [ ] sc11d2l3 (790)

   The [P] indicates the current path.

Switch configurations

The 200/300 Series storage systems tested successfully in a fabric connected to the SGI Origin 2000. The 200/300 Series and SGI box both did full fabric log-ins. The following were the levels of various software tested, which continues with newer firmware, applications, and management software.

1. firmware – 04.01.00.03, 04.01.02.10, 04.01.02.20, 05.01.01.00, 05.36.00.00, and 05.36.01.00

2. NVSRAM - NV4774HET801022, NV2772HET801023, N4884-500855-001, N4774-500855-001 and -536855-001

3. IRIX 6.5.14 feature with SGI Patch 4417

4. 4116 switch, and a 2 Gb switch (currently under test)

5. `/var/sysgen/master.d/qlfc` file variable changed from 0 to 1 (point to point)

   qlfc_use_connection_mode:

   > 3 - point to point preferred, then loop

   > 2 - loop preferred, then point to point

   > 1 - point to point only

   > 0 - loop mode

   qlfc_use_connection_mode = 1;

Note: Engineering tests tried using option 3, but the Qlogic card would intermittently end up in loop mode.

6. `/var/sysgen/master.d/qlfc` file variable changed from 240 to 20 (timeout same as direct connect)

   qlfc_target_missing_timeout = 20;

Configuring XLV volumes and failover work the same way but have the World Wide Name/Lun_Slice/Hwpath constructed in the /dev/dsk directory structures Consult man pages for xlv_make, failover, scsifo, etc. for syntax.

The hinv output looks like:

Two 195 MHZ IP27 Processors

CPU: MIPS R10000 Processor Chip Revision: 2.6

FPU: MIPS R10010 Floating Point Chip Revision: 2.6

CPU 0 at Module 1/Slot 1/Slice A: 195 Mhz MIPS R10000 Processor Chip (enabled)

  Processor revision: 2.6. Scache: Size 4 MB Speed 130 Mhz   Tap 0x9

CPU 1 at Module 1/Slot 1/Slice B: 195 Mhz MIPS R10000 Processor Chip (enabled)

  Processor revision: 2.6. Scache: Size 4 MB Speed 130 Mhz   Tap 0x9

Main memory size: 256 Mbytes

Instruction cache size: 32 Kbytes

Data cache size: 32 Kbytes

Secondary unified instruction/data cache size: 4 Mbytes

Memory at Module 1/Slot 1: 256 MB (enabled)

  Bank 0 contains 128 MB (Standard) DIMMS (enabled)

  Bank 1 contains 128 MB (Standard) DIMMS (enabled)

Integral SCSI controller 0: Version QL1040B (rev. 2), single ended

  Disk drive: unit 1 on SCSI controller 0 (unit 1)

  CDROM: unit 6 on SCSI controller 0

Integral SCSI controller 1: Version QL1040B (rev. 2), single ended

Integral SCSI controller 11: Version Fibre Channel QL2310

  Fabric Disk: node 200400a0b80f4617 port 200500a0b80f4618 lun 0 on SCSI controller 11 (unit 254) (primary path)

  Fabric Disk: node 200400a0b80f4617 port 200500a0b80f4618 lun 1 on SCSI controller 11 (unit 254) (primary path)

  Fabric Disk: node 200400a0b80f4617 port 200500a0b80f4618 lun 2 on SCSI controller 11 (unit 254) (alternate path) DOWN

  Fabric Disk: node 200400a0b80f4617 port 200500a0b80f4618 lun 3 on SCSI controller 11 (unit 254) (alternate path) DOWN

  Fabric Disk: node 200400a0b80f4617 port 200500a0b80f4618 lun 5 on SCSI controller 11 (unit 254)

  Fabric Disk: node 200400a0b80f4617 port 200500a0b80f4618 lun 31 on SCSI controller 11 (unit 254)

Integral SCSI controller 12: Version Fibre Channel QL2310

  Fabric Disk: node 200400a0b80f4617 port 1 lun 0 on SCSI controller 12 (unit 254) (alternate path) DOWN

  Fabric Disk: node 200400a0b80f4617 port 1 lun 1 on SCSI controller 12 (unit 254) (alternate path) DOWN

  Fabric Disk: node 200400a0b80f4617 port 1 lun 2 on SCSI controller 12 (unit 254) (primary path)

  Fabric Disk: node 200400a0b80f4617 port 1 lun 3 on SCSI controller 12 (unit 254) (primary path)

  Fabric Disk: node 200400a0b80f4617 port 1 lun 5 on SCSI controller 12 (unit 254)

Fabric Disk: node 200400a0b80f4617 port 1 lun 31 on SCSI controller 12 (unit 254)

Integral SCSI controller 4: Version QL1040B (rev. 2), differential

Integral SCSI controller 6: Version QL1040B (rev. 2), differential

Integral SCSI controller 7: Version QL1040B (rev. 2), differential

Integral SCSI controller 5: Version QL1040B (rev. 2), differential

Integral SCSI controller 8: Version Fibre Channel AIC-1160, revision 1

Integral SCSI controller 9: Version Fibre Channel AIC-1160, revision 1

IOC3 serial port: tty1

IOC3 serial port: tty2

IOC3 serial port: tty3

IOC3 serial port: tty4

IOC3 serial port: tty5

IOC3 serial port: tty6

IOC3 serial port: tty7

IOC3 serial port: tty8

Integral Fast Ethernet: ef0, version 1, module 1, slot io1, pci 2

Fast Ethernet: ef1, version 1, module 1, slot io4, pci 0

Fast Ethernet: ef2, version 1, module 1, slot io4, pci 1

Fast Ethernet: ef3, version 1, module 1, slot io4, pci 2

Fast Ethernet: ef4, version 1, module 1, slot io4, pci 3

Origin BASEIO board, module 1 slot 1: Revision 3

  PCI Adapter ID (vendor 4265, device 3) pci slot 2

  PCI Adapter ID (vendor 4215, device 4128) pci slot 0

  PCI Adapter ID (vendor 4215, device 4128) pci slot 1

Origin MENET board, module 1 slot 4: Revision 3

  PCI Adapter ID (vendor 4265, device 3) pci slot 0

  PCI Adapter ID (vendor 4265, device 3) pci slot 1

  PCI Adapter ID (vendor 4265, device 3) pci slot 2

  PCI Adapter ID (vendor 4265, device 3) pci slot 3

Origin PCI XIO board, module 1 slot 2: Revision 4

PCI Adapter ID (vendor 4215, device 8960) pci slot 1

PCI Adapter ID (vendor 4215, device 8960) pci slot 2

Origin MSCSI board, module 1 slot 5: Revision 3

PCI Adapter ID (vendor 4215, device 4128) pci slot 0

PCI Adapter ID (vendor 4215, device 4128) pci slot 1

PCI Adapter ID (vendor 4215, device 4128) pci slot 2

PCI Adapter ID (vendor 4215, device 4128) pci slot 3

Origin FIBRE CHANNEL board, module 1 slot 3: Revision 3

PCI Adapter ID (vendor 36868, device 4448) pci slot 0

PCI Adapter ID (vendor 36868, device 4448) pci slot 1

IOC3 external interrupts: 1

HUB in Module 1/Slot 1: Revision 5 Speed 97.50 Mhz (enabled)

IP27prom in Module 1/Slot n1: Revision 6.80

IO6prom on Global Master Baseio in Module 1/Slot io1: Revision 6.80

fx –x –d

The customer will need to partition and label the disks using the fx command. There are 4 physical LUNs in the previous example and 8 paths: LUN 1 is used here for an example of the required syntax. Formatting the volumes in a switched environment requires the use of the –d flag (see manual page for fx utility); and it would look similar to the following example.

```
fx –x –d /dev/rdsk/200800a0b807b3ae/lun1vol/c11p200500a0b80f4618
```

This sets the sequence of events allowing customers to partition their disk; and also establish the primary path in the failover configuration which is addressed later through the failover.conf.

The alternate path will not be assigned through the fx utility. Similar commands need to be performed for each of the customer's LUNs.

The failover.conf entries look like (from /dev/scsi paths):

```
#ident $Revision: 1.1 $
#
#     This is the configuration file for table driven failover support.
#     The entries within this file consist of a single line per
```

```
#     failover grouping.  These lines describe up to four paths to
#     each device.  Obviously, these paths should reference the
#     same device!  Some sanity checking is performed, but the
#     software is not foolproof in this area. (If it was foolproof,
#     the configuration file would be unnecessary!)  It is recommended
#     that all paths to a device be listed.
#
#     The format of each line consists of a group name followed by
#     up to four device names from the "/dev/scsi" directory.  If
#     it's not in "/dev/scsi", it cannot be configured.  But, this
#     implementation assumes that if a device is not present, its
#     path has failed.  Consequently, no error messages will be
#     generated for missing paths.  This means that you may have a
#     a failover group that consists of a single primary path with
#     no alternates.
#
#     Additionally, there is a configuration directive available that
#     will cause the program to emit debug information.  Placing
#     "#verbose" at the start of a line, without the quotes, will
#     enable the debug output.  The debug output is displayed for
#     all subsequent configuration directives.  Once enabled, the
#     debug output cannot be disabled.
#
#     Lines that begin with a "#" are considered comment lines.
#     A "#" anywhere within a line signals the beginning of a comment.
#     White space must separate the last parameter of a line and the #.
#     Blank lines are also considered a comment.
#
#     --> Lines that begin with 'sc' will be skipped as it's likely a
#     --> configuration error.
#
#     Sample configuration.  (Remember to omit the "#" the start of
#     the line.)
#
#
#  # Name     Pri path | alt path | alt path | alt path
#
#     GroupA  sc6d5ol0   sc7d5ol0   sc8d5ol0   sc9d5ol0
#     GroupB  sc6d5ll0   sc7d5ll0   sc8d5ll0   sc9d5ll0       # Some comment
#     GroupC  sc6d5ol0   sc7d5ol0   sc8d5ol0   sc9d5ol0
#
#     One word of caution.  The switch to an alternate path is
#     choreographed by XLV.  XLV is a requirement for failover to
#     function.
#
#     Notes:
#
#     It is not possible to change the primary device of a group after
#     the initial configuration by reordering the entries on the line
#     and running /etc/init.d/failover start.  To change the primary,
#     scsifo may be used to switch to the next available path.
#
#     It is not possible to remove the primary of a group by deleting
#     the entry from the group and running /etc/init.d/failover start.
#     A reboot is necessary.
#
#     To remove the "DOWN" displayed by hinv for the failed path,
#     fix the path and reprobe the bus using scsiha -p #.
#
#verbose
```

```
#verbose
#disable_target_lun_check

vol0    200400a0b80f4617/lun0/c11p200500a0b80f4618    200400a0b80f4617/
lun0/c12p1
vol2    200400a0b80f4617/lun2/c12p1   200400a0b80f4617/lun2/
c11p200500a0b80f4618
vol1    200400a0b80f4617/lun1/c11p200500a0b80f4618    200400a0b80f4617/
lun1/c12p1
vol3    200400a0b80f4617/lun3/c12p1   200400a0b80f4617/lun3/
c11p200500a0b80f4618
```

Obtaining HBA World Wide Names for Partitioning

To obtain the WWN of the host adapters for use in partitioning, run the following commands:

1. Run hinv and note the controller numbers of the HBAs:

    Two 195 MHZ IP27 Processors

    CPU: MIPS R10000 Processor Chip Revision: 2.6

    FPU: MIPS R10010 Floating Point Chip Revision: 0.0

    CPU 0 at Module 1/Slot 1/Slice A: 195 Mhz MIPS R10000 Processor Chip (enabled)

      Processor revision: 2.6. Secondary cache: Size 4 MB Speed 130 Mhz

    CPU 1 at Module 1/Slot 1/Slice B: 195 Mhz MIPS R10000 Processor Chip (enabled)

      Processor revision: 2.6. Secondary cache: Size 4 MB Speed 130 Mhz

    Main memory size: 256 Mbytes

    Instruction cache size: 32 Kbytes

    Data cache size: 32 Kbytes

    Secondary unified instruction/data cache size: 4 Mbytes

    Memory at Module 1/Slot 1: 256 MB (enabled)

      Bank 0 contains 128 MB (Standard) DIMMS (enabled)

      Bank 1 contains 128 MB (Standard) DIMMS (enabled)

    Integral SCSI controller 0: Version QL1040B (rev. 2), single ended

      Disk drive: unit 1 on SCSI controller 0 (unit 1)

      CDROM: unit 6 on SCSI controller 0

    Integral SCSI controller 1: Version QL1040B (rev. 2), single ended

Integral SCSI controller 10: Version Fibre Channel QL2200A

  Disk drive: unit 5 on SCSI controller 10 (unit 5)

  Disk drive: unit 5, lun 1 on SCSI controller 10 (unit 5)

  Disk drive: unit 5, lun 2 on SCSI controller 10 (unit 5)

  Disk drive: unit 5, lun 3 on SCSI controller 10 (unit 5)

Integral SCSI controller 11: Version Fibre Channel QL2200A

  Disk drive: unit 2 on SCSI controller 11 (unit 2)

  Disk drive: unit 2, lun 1 on SCSI controller 11 (unit 2)

  Disk drive: unit 2, lun 2 on SCSI controller 11 (unit 2)

  Disk drive: unit 2, lun 3 on SCSI controller 11 (unit 2)

Integral SCSI controller 4: Version QL1040B (rev. 2), differential

Integral SCSI controller 5: Version QL1040B (rev. 2), differential

In this example, the Qlogic cards (QLA2200A) are at controller numbers 10 and 11.

2. Run `scsiha -w n` - where n is the controller number. Example:

For the example above, you need WWNs from controllers 10 and 11:

```
scsiha -w 10
```

10 Portname: 210000e08b014f63

```
scsiha -w 11
```

11 Portname: 210000e08b015262

NEED MORE INFORMATION?

www.storagetek.com

ABOUT STORAGETEK

Storage Technology Corporation (NYSE: STK) is a $2 billion global company that enables businesses, through its information lifecycle management strategy, to align the cost of storage with the value of information. The company's innovative storage solutions manage the complexity and growth of information, lower costs, improve efficiency and protect investments. For more information, visit www.storagetek.com, or call 1.800.275.4785 or 01.303.673.2800.

WORLD HEADQUARTERS

Storage Technology Corporation
One StorageTek Drive
Louisville, Colorado 80028 USA
1.800.525.0369