Sun StorEdge™ Fast Write Cache 2.0 Configuration Guide

THE NETWORK IS THE COMPUTER™

Sun Microsystems, Inc.
901 San Antonio Road
Palo Alto, CA 94303-4900 USA
650 960-1300 Fax 650 969-9131

Part No. 806-4383-10
February 2000, Revision A

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

Please
Recycle

Sun StorEdge Fast Write Cache Configuration Guide

Introduction

Sun StorEdge™ Fast Write Cache is a host-based write accelerator for A5X00 storage systems. It improves performance for transaction processing and delivers faster response times to user requests for data by reducing the frequency of disk I/O accesses. Writes are cached in non-volatile memory, then the cached data is destaged to disk. Fast Write Cache is installed on Solaris™ servers and consists of NVRAM (non-volatile memory) boards used as cache memory and storage cache management software.

Note – Fast Write Cache is implemented using a pair of SBus or PCI NVRAM cards and storage cache management software on your Solaris server. One pair of cards is supported per system, excepting the Enterprise 10000, where one pair is supported per domain.

Contents

This document contains the following information:

- Configuration Considerations
- Performance
- Sample Configurations

Installation Information

Refer to the *Sun StorEdge Fast Write Cache 2.0 Installation Guide* for the following installation-related information:

- System Requirements
- Qualified Platforms
- Software Compatibility
- Limitations

Related Documentation

TABLE P-1 Related Documentation

Application	Title	Part Number
`man` Pages	`fwcadm(1FWC)` `svadm(1SV)`	N/A
Release Notes	*Sun StorEdge Fast Write Cache 2.0 Release Notes*	806-0476-10
Installation	*Sun StorEdge Fast Write Cache 2.0 Installation Guide*	806-4405-10
User	*Sun StorEdge Fast Write Cache 2.0 System Administrator's Guide*	806-2064-10
White Papers	*Sun StorEdge Fast Write Cache Technical White Paper*	
Best Practice Guides	*Improving OLTP Performance on the Sun StorEdge A5000*	

Applications

The A5X00 is a performance leader in a variety of applications including data warehousing, decision support and imaging. Sun StorEdge Fast Write Cache enables Sun Microsystems™ to extend the performance leadership of the A5X00 to additional applications such as OLTP and file serving.

Performance improvements have been measured between 0-40X depending on the size of transfers, the number of threads, and other factors. The highest rate of improvement has been observed for sequential writes, when the cache coalesces the small writes into larger writes. For RAID 5, these writes become full strip writes, thus eliminating the read-modify-write scenario.

Fast Write Cache allows the system administrator to choose which volumes get cached and which do not. The system administrator should cache:

- NFS™ data
- Data used by OLTP applications

The system administrator should not cache:

- Data used by Decision Support applications
- Read-Only volumes

Architecture

Sun StorEdge Fast Write Cache is implemented as a UNIX® device driver, using nonvolatile memory to cache write requests. As the cache fills, older data is written asynchronously to the real disk. Fast Write Cache works as a layer between other disk drivers and the rest of the UNIX kernel. Stubs replace the original driver's entry points in the device switch tables. Whenever Fast Write Cache performs actual I/O (for example, when its cache must be destaged), it uses the real device driver routines.

Fast Write Cache sits above the Volume Manager, allowing it to work with any type of volume, whether it is RAID 5, RAID 1, or RAID 0+1.

Hardware Components

Fast Write Cache Release 2.0 is comprised of dual redundant PCI or SBus NVRAM cards. These cards have 32MBytes of capacity. The cards are equipped with ECC single bit error detection and correction and ECC double bit error detection. Should a double bit error occur, software will take the card off-line, destage the cache, and put the cache into write-through mode. During normal I/O operations, all accesses are writes. During recoveries (after an unclean shutdown), blocks are read from NVRAM.

SBus NVRAM Card

Each card is single-wide, with two mezzanine cards and three lithium batteries. The shelf life of the batteries is five years. The batteries are capable of data retention for up to 18 months. Each battery is enabled by a jumper on the card.

The measured write performance of the card is 108 MBytes/S.

The NVRAM card contains a single LED which is powered by the system, not the on-board batteries.

The light is on when:

- Cache is enabled
- There was an unclean shutdown while the cache was enabled

The light is off when

- Cache is disabled
- Card is not installed
- Card is installed, but the system power is off

PCI NVRAM Card

The PCI NVRAM card is full-length, with 64MBytes of capacity, and four batteries. The shelf-life of the batteries is ten years; their data retention is approximately 34 months. The batteries are activated by a single enable/disable switch on the card. The data transfer rate is up to 90 MBytes per second.

Software Components

Fast Write Cache is comprised of the following software components:

- Storage Volume Driver (`SUNWspsv`)

- Storage Cache Manager (SUNWscm)
- NVRAM driver (SUNWnvm)
- Diagnostic (SUNWvtsnp)
- Uniform status reporting (SUNWspuni)

Storage Volume Driver (SUNWspsv)

Storage Volume Driver is a layered driver that intercepts I/O calls to the cached devices. It then copies data to and from the Solaris operating environment buffers to buffers internal to the cache.

Storage Cache Manager (SUNWscm)

The Storage Cache Manager contains the cache software and performs the following functions:

- Coalescing
- Destaging
- Switching into and out of write-through mode
- Manager-pinned data - data that is in the cache and cannot be destaged to disk (in the case of a disk failure, for example)

NVRAM Driver (SUNWnvm)

SUNWnvm contains the software necessary to support the NVRAM cards and handles the mirroring between the redundant pair of NVRAM cards. It also contains software that monitors the state of the memory on the cards, batteries, and general health of the cards. If it detects a card failure, it notifies the Storage Cache Manager which will then switch to write-through mode.

Diagnostic (SUNWvtsnp)

SUNWvtsnp contains diagnostics for the NVRAM cards.

Uniform status reporting (SUNWspuni)

This module provides error reporting services for the data services.

Configuration Considerations

To Cache or Not to Cache

The Fast Write Cache allows the system administrator to choose which volumes get cached and which volumes do not. This provides the best of both worlds; direct access to high-speed disks when needed and caching when needed.

With the ability to dynamically insert and remove volumes from the cache, the system administrator can cache volumes when caching make sense and go directly to the disk when that is appropriate.

When to Cache

- Cache the data of applications that execute small, sequential writes
  - Database Logs
  - OLTP data
  - File system data
- Cache RAID 5 volumes

When Not to Cache

Do not cache the following entities:

- Underlying volume manager devices (for example, VxVM log devices), as this can result in unpredictable behavior.
- Root (`/`) and `/usr` file systems (or any file systems that come up before Fast Write Cache does); in order to recover disk data stored in an NVRAM card after a system crash, the data must be restored to disk before file systems are mounted or applications are written to raw volumes.
- The swap partition.
- Data of applications that perform large, asynchronous writes
  - Example: Decision support applications
- Read-only volumes

Raw or Cached Volumes

Volumes can be inserted into the cache as *raw* or *cached*. When a volume is inserted as *cached*, the I/O finishes when the data is stored in the NVRAM memory. If *raw* is specified, the I/O completion acknowledgment does not occur until the data is on the physical media. This enables volumes to use the cache interface but not be cached. This is required if the system administrator wants to make the volume available for data services (such as Instant Image), but does not want to cache the volume, or has not installed the NVRAM cards. In the case of the NVRAM cards not being installed or functioning, the cache operates in write-through mode. That is, the I/O completion acknowledgment is not given to the application until the data is stored on the physical disks.

Performance

Applications executing sequential writes gain the biggest performance improvement using Fast Write Cache. The following table illustrates the performance improvement of single-threaded sequential writes. The table shows the differences between two physical A5000 7200RPM drives, one cached and one not cached.

TABLE 0-1 Sequential Write Performance Comparison

Sequential I/O	Throughput	IOPS	Response Time
2KB Noncached	187.93KB/s	94.04	11ms
2KB Cached	9684.09KB/s	4843.98	(1) 0ms
8KB Noncached	663.06KB/s	82.92	12ms
8KB Cached	10550.14KB/s	1319.34	1ms
16KB Noncached	1169.21KB/s	73.11	14ms
16KB Cached	10567.28KB/s	660.79	1ms
256KB Noncached	6320.23KB/s	24.71	40ms
256KB Cached	10822.59KB/s	43.30	24ms

(1) Response Time is less than 500 microseconds, rounded to 0.000 seconds, which when converted to milliseconds is 0.

Please refer to the *Fast Write Cache Technical White Paper* for more detailed performance information on this product.

Sample Configurations

TABLE 1

Application	Hardware Configuration	Workload Mix	I/O Size	RAID Configuration	Caching Recommendation
OLTP	E3000 A5000	75% Reads 25% Writes	2 KB	RAID 1	Cache Redo Logs
	E4500 A5200	60% Reads 40% Writes	8KB	RAID 1	Cache volumes to which writes occur
Decision Support	E6500 A5200	90% Reads 10% Writes	64KB	RAID 5	Do not cache