SUN BLUEPRINTS APPROVED

Sun StorEdge™ T3 Dual Storage Array - Part 1

Installation, Planning, and Design

By Mark Garner - Enterprise Engineering

Sun BluePrints™ OnLine - February 2001

Sun microsystems
ENTERPRISE ENGINEERING

http://www.sun.com/blueprints

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

Part No.: 806-6641-10
Revision 01, 02/06/01
Edition: February 2001

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.

CETTE PUBLICATION EST FOURNIE "EN L'ETAT" ET AUCUNE GARANTIE, EXPRESSE OU IMPLICITE, N'EST ACCORDEE, Y COMPRIS DES GARANTIES CONCERNANT LA VALEUR MARCHANDE, L'APTITUDE DE LA PUBLICATION A REPONDRE A UNE UTILISATION PARTICULIERE, OU LE FAIT QU'ELLE NE SOIT PAS CONTREFAISANTE DE PRODUIT DE TIERS. CE DENI DE GARANTIE NE S'APPLIQUERAIT PAS, DANS LA MESURE OU IL SERAIT TENU JURIDIQUEMENT NUL ET NON AVENU.

Sun StorEdge™ T3
Dual Storage Array - Part 1
Installation, Planning, and Design

The intended audience for this article is system administrators and any other parties responsible for the design and configuration of Sun StorEdge™ T3 arrays.

This is the first of three articles that will provide a roadmap for the configuration of the Sun StorEdge T3 storage array partner group by discussing the following points:

- Installation planning and design—Part 1
- Configuration—Part 2
- Basic Management—Part 3

Information presented in this article is discussed in greater detail in the Sun StorEdge T3 array, VERITAS Volume Manager, and Sun™ Management Center documentation. Available at http://docs.sun.com/ and http://www.veritas.com/.

The following best practices are featured in this series of articles:

- Layout of storage (manageability)
- Using VERITAS Volume Manager (VxVM) (maintainability)
- Wide Thin striping (performance)
- Basic systems management and monitoring (availability)

The third article will also look at how to automate configurations using Expect scripts.

Installation Planning and Design

The sections that follow present planning and design requirements for the installation of a T3 array partner group. The installation planning and design specifications are summarized in appendix A.

Physical Configuration

The T3 storage array features a hardware Redundant Array of Independent Disks (RAID) controller. There are currently two supported T3 array configurations:

- Single Array — Sun StorEdge T3 Array for the Workgroup

- Dual arrays (fully redundant) — Sun StorEdge T3 Array for the Enterprise

A single T3 array is redundant except for the controller card. In a high availability environment where full redundancy is required, two T3 arrays with RAID controllers must be employed.

Power

Two power feeds are required for each T3 array to increase resiliency. For a high availability configuration, each supply should be powered from a separate source.

Note – New equipment should be powered up during a planned quiet period to minimize possible disruptions.

T3 Array Partner Group

To provide RAID controller resilience, two T3 arrays must be connected together. Two interconnect cables are used to achieve this, and the arrays become known as a partner group or pair. They function as a master/alternate master configuration, meaning, if the master controller fails, then the alternate master will take over its functions. Each controller is able to operate both arrays, but under normal operation each will handle the functions of its own array. In all other respects, once partnered, the arrays behave as if they were a single array.

IP Address

The T3 controller has a 10 BaseT interface used for configuration and management. In a partner group, only the IP address of the master array is accessible, therefore only a single network address is required. The following information in Table 1 should be defined (example addresses are shown):

TABLE 1 IP/Netmask/Gateway Information

T3 IP Address	192.168.49.183
Netmask	255.255.255.0
Gateway IP Address	192.168.49.254

Note – If the network controller of the master array fails, the network address will failover to the alternate master. However, if the network cable is disconnected from the master array, the network address does not failover to the alternate master.

Dual Array Volume Layout

Architecture

The T3 array uses a switched-loop architecture which employs two fibre channel connections for each disk drive within the array. The switched-loop architecture enables the drive interface to be split into multiple, independent loops, in this instance two. Each RAID controller uses one loop, but has access to the others for resilience.

If two T3 arrays are connected as a partner group, the pair of interconnect cables carry redundant fibre channel loops and control buses between arrays. This enables either RAID controller to access any disk in the partner group.

To help ensure resilience, there should be two connections between the host and the T3 array partner group; one from each RAID controller to different I/O boards on the host. This further eliminates a single point of failure (SPOF), as illustrated in FIGURE 1.

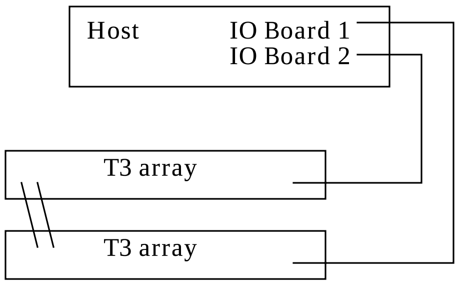

FIGURE 1 Host to Array Fibre Channel Connections

The T3 array enables disks to be grouped together—which are termed volumes. The volumes appear as Logical Unit Numbers (LUNs) to the host, for example, c0t0d0 and c0t0d1—that is, controller 0, target 0, LUNs 0 and 1.

A T3 array can be configured with a maximum of two volumes. That is, for a partner group a maximum of four volumes can be created. A volume is restricted to a minimum of two disks.

One hot spare disk can be defined for each array—it must always be disk number 9. The hot spare will secure volumes on the same array, but not on the partner array, and will be substituted for a failed disk in RAID 1 and RAID 5 configurations. RAID configurations will be described later in this article. A hot spare must be defined where availability is the most important design requirement.

Dual host-array connections require the use of Veritas Volume Manager (VxVM) to provide resilience. The T3 array partner group will present each volume to each host-array connection. To illustrate this point, if a single volume is defined on a T3 array partner group, it will be called volume 0 (v0) and will be defined as LUN 0. The host will define this as LUN 0 (d0) on both host adapters, i.e., c1t0d0 & c2t0d0.

VxVM provides Dynamic MultiPathing (DMP) a feature that will manage the failover of disk volumes between the host adapters.

RAID

The T3 array will support three RAID levels 0, 5 and 1+0:

RAID 0 (or striping) - blocks of data are striped across disks (disks 2-9 on a T3 array). A read or write is split into multiple parallel operations, thereby increasing the data throughput accordingly. RAID 0 is sometimes referred to as the suicide stripe—a single disk failure will destroy all data. Unless host based mirroring is employed (where the host mirrors the data to another set of disks), RAID 0 should be avoided.

RAID 1+0 or striped mirrors - disks are divided into mirror pairs or groups with blocks of data being striped across the mirror groups. A read or write is split into multiple parallel operations, thereby adding resilience by writing the data to more than one disk. RAID 1+0 is able to suffer multiple disk failures, providing that all the disks belonging to a mirror group do not fail. A hot spare can be defined, and will take over from the failure of another disk. This configuration can be used with single T3 array and host based mirroring, or a T3 array partner group. In both configurations, the data must be mirrored between arrays for increased resilience.

RAID 5 - data is striped across disks, additionally, a parity calculation is performed on the data and is written to disk. The parity data is striped across all disks to avoid a bottleneck. RAID 5 offers two potential advantages; over RAID 0, the data is secured; over RAID 1+0, securing the data requires only a one disk overhead. The major disadvantage with RAID 5 is that the parity calculation imposes an overhead large enough to make host based RAID 5 unrealistic for most applications.

The good news is that when using hardware RAID, the controller performs the parity calculation, thereby, potentially negating the disadvantage. In the case of the T3, the performance of a RAID 5 is comparable to RAID 1 for most workloads (except small random write operations where RAID 1 performance is approximately 20% better).

Where resilience is the key requirement, the recommended configuration is to use a T3 array partner group and hardware RAID 5. Two volumes should be configured (one per array) as follows:

- 7 + 1 disks (RAID 5), 1 disk (hot spare) on the upper array
- 7 + 1 disks (RAID 5), 1 disk (hot spare) on the lower array

This configuration maximizes the storage space without sacrificing resilience, and off loads the overhead of RAID management from the host to the array.

Stripe Unit Size

The stripe unit size on a T3 array can be set to 16, 32, or 64 Kbytes by using the `sys blocksize n` command when logged in to the T3 array (where `n` will be 16, 32, or 64 Kbytes). The stripe unit size can only be changed if no volumes are defined. It is suggested that you use the default value of 64 Kbytes (unless the behavior of the application is known to require a smaller value).

Volume Manager

In a high availability environment (or where the application requires relatively small file systems), a volume manager should be employed. The features of a volume manager enhance manageability, maintainability, and therefore, system availability as described in this section.

The Sun StorEdge T3 array for the Enterprise product is supplied with Veritas Volume Manager (VxVM) media and licence. This is done for the following reasons:

- The disk space that is created with a 2 x (7 + 1) RAID 5 configuration is about 224 Gbytes (using 18.2 Gbyte disks). This size could be difficult to manage as two (separate) 112 Gbyte file systems.

Note – The T3 can be ordered with 36.4 or 73 Gbyte drives.

- The VxVM product enables the disks to be divided into sub-disks—each of which can be designated to perform a specific function. For example, index, data, logs, and archive in a relational database application. Because the sub-disks are spread over eight physical disks, the process of balancing data I/O across spindles is inherent to the design.
- The VxVM product should allow volumes to be *dynamically grown* in size.
- The dynamic multipathing (DMP) feature of VxVM provides resilience for the host-to-array connections—this should allow I/O operations to be transparently continued on the surviving Fibre Channel link in the event that one of the two links fail.

Systems Management

Sun StorEdge T3 arrays are supplied with Sun StorEdge Component Manager software. This software should be installed on an appropriate host, for example, a systems management or systems administration server. The Sun StorEdge Component Manager software can be used to monitor and manage the T3 arrays and can route alerts to e-mail. A management host, where the software is to be installed, must be chosen prior to installation—the destination address of this host is where the alerts are sent.

The T3 arrays can be configured to forward simple network management protocol (SNMP) traps—if you use this feature, you must determine prior to configuration a destination host address that will receive the traps.

Summary

This article has presented an overview of the pre-requisites for a T3 array partner group installation and covered the requirements for:

- Power and Cooling.
- Network connections.
- Network address.
- Host-Array connections.
- Storage Volume layout.
- Volume Manager software.
- Systems Management host and software.

The second article in the series will describe how to install and configure a T3 array partner pair to the specification defined in this article.

Appendix A - Installation planning and design specifications

The pre-requisite information required to install the T3 array is as follows:

- **Physical Configuration**

  Each Sun StorEdge T3 array has the following physical characteristics:

  - Height: 5.25
  - Width: 17.5
  - Depth: 18.5 + 4 inches front and back for cooling
  - Weight: 67lb
- **Power**
  - 4 x 100 - 120 or 200 - 240VAC (450W) electrical supplies from independent sources
  - 2 x 1540 BTU/hr. cooling
- **Physical Connections**
  - 2 x 10 BaseT ethernet connection between the T3 array and the local area network.
  - 2 x Fibre Channel (FC-AL) connection between the T3 array and the host.
- **IP Addresses**
  - One TCP/IP address with subnet mask, hostname and gateway information.
- **Volume Layout**
  - 2 x 7+1 (Raid 5) volume with one hot spare for performance (wide thin striping) and availability.
- **Volume Manager**
  - Install VERITAS Volume Manager for maintainability and resilience.
- **Systems Management**
  - Install Sun StorEdge Component Manager software for manageability, or use syslog and redirect events to a remote or log host.

Author's Bio: Mark Garner

Mark Garner is a Systems Engineer for Enterprise Engineering at Sun Microsystems. Prior to joining Sun, Mark spent three years as a Systems Architect specializing in Email and Office Automation architectures and before this over eight years in systems administration. While at Sun Mark has focused on the design and implementation of mission critical business and Internet application's infrastructure principally to provide for high availability and scalability.