Scalable Enterprise Storage Solutions

For Systems Based on the Intel® Pentium® III Xeon™ Processor

Revision History

Date	Revision	Modifications
August, 1999	1.0	Initial release.

Disclaimers

The information contained in this document is provided for informational purposes only and represents the current view of Intel Corporation (Intel) and its contributors ("Contributors") LSI Logic and CLARiiON Advanced Storage Division of Data General Corporation, as of the date of publication. Intel makes no commitment to update the information contained in this document, and Intel reserves the right to make changes at any time, without notice.

THIS DOCUMENT, "SCALABLE ENTERPRISE STORAGE SOLUTIONS FOR SYSTEMS BASED ON THE INTEL PENTIUM® III XEON™ PROCESSOR," IS PROVIDED "AS IS." NEITHER INTEL NOR THE CONTRIBUTORS MAKE ANY REPRESENTATIONS OF ANY KIND WITH RESPECT TO PRODUCTS REFERENCED HEREIN, WHETHER SUCH PRODUCTS ARE THOSE OF INTEL, THE CONTRIBUTORS OR THIRD PARTIES. INTEL AND ITS CONTRIBUTORS EXPRESSLY DISCLAIM ANY AND ALL WARRANTIES, IMPLIED OR EXPRESS, INCLUDING WITHOUT LIMITATION, ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR ANY PARTICULAR PURPOSE, NON-INFRINGEMENT, AND ANY WARRANTY ARISING OUT OF THE INFORMATION CONTAINED HEREIN, INCLUDING WITHOUT LIMITATION, ANY PRODUCTS, SPECIFICATIONS, OR OTHER MATERIALS REFERENCED HEREIN. INTEL AND ITS CONTRIBUTORS DO NOT WARRANT THAT THIS DOCUMENT IS FREE FROM ERRORS, OR THAT ANY PRODUCTS OR OTHER TECHNOLOGY DEVELOPED IN CONFORMANCE WITH THIS DOCUMENT WILL PERFORM IN THE INTENDED MANNER, OR WILL BE FREE FROM INFRINGEMENT OF THIRD PARTY PROPRIETARY RIGHTS, AND INTEL AND ITS CONTRIBUTORS DISCLAIM ALL LIABILITIES THEREFOR.

INTEL AND ITS CONTRIBUTORS DO NOT WARRANT THAT ANY PRODUCT REFERENCED HEREIN OR ANY PRODUCT OR TECHNOLOGY DEVELOPED IN RELIANCE UPON THIS DOCUMENT, IN WHOLE OR IN PART, WILL BE SUFFICIENT, ACCURATE, RELIABLE, COMPLETE, FREE FROM DEFECTS OR SAFE FOR ITS INTENDED PURPOSE, AND HEREBY DISCLAIM ALL LIABILITIES THEREFOR. ANY PERSON MAKING, USING OR SELLING SUCH PRODUCT OR TECHNOLOGY DOES SO AT HIS OR HER OWN RISK.

Licenses may be required. Intel, its contributors and others may have patents or pending patent applications, trademarks, copyrights or other intellectual proprietary rights covering subject matter contained or described in this document. No license, express, implied, by estoppel or otherwise, to any intellectual property rights of Intel or any other party is granted herein. It is your responsibility to seek licenses for such intellectual property rights from Intel and others where appropriate.

Limited License Grant. Intel hereby grants you a limited copyright license to download and copy this document for your use and internal distribution only. You may not distribute this document externally, in whole or in part, to any other person or entity.

LIMITED LIABILITY. IN NO EVENT SHALL INTEL OR ITS CONTRIBUTORS HAVE ANY LIABILITY TO YOU OR TO ANY OTHER THIRD PARTY, FOR ANY LOST PROFITS, LOST DATA, LOSS OF USE OR COSTS OF PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES, OR FOR ANY DIRECT, INDIRECT, SPECIAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF YOUR USE OF THIS DOCUMENT OR RELIANCE UPON THE INFORMATION CONTAINED HEREIN, UNDER ANY CAUSE OF ACTION OR THEORY OF LIABILITY, AND IRRESPECTIVE OF WHETHER INTEL HAS ADVANCE NOTICE OF THE POSSIBILITY OF SUCH DAMAGES. THESE LIMITATIONS SHALL APPLY NOTWITHSTANDING THE FAILURE OF THE ESSENTIAL PURPOSE OF ANY LIMITED REMEDY.

The Intel® AC450NX MP Server System and the Intel® OCPRF100 MP Server System may contain design defects or errors known as errata that may cause the product(s) to deviate from published specifications. Current characterized errata are available upon request from Intel.

** PCI Hot-plug functionality is based upon a licensed design from Compaq Corporation.

Conventions and Terminology

This document uses the following terms and abbreviations:

Term	Definition
AKA	Also Known As
CPU	Central Processing Unit
DB2*	IBM*'s Database
DAE	Disk Array Enclosure
DPE	Disk Processor Enclosure
DMI	Desktop Management Interface
Fail-over	Transfer functionality to another system
FC	Fibre Channel
FC-AL	Fibre Channel Arbitrated Loop
FRU	Field Replaceable Unit
HBA	Host Bus Adapter
HDD	Hard Disk Drive
Hot-plug	Plugging the adapters while system power is ON
IA	Intel® Architecture
ISC	Intel® Server Control
JBOD	Just a Bunch Of Disks (disks that are directly accessed as opposed to RAID)
LAN	Local Area Network
LCC	Link Controller Card
Like for Like	PCI Hot-plug replacement of same type of HBA
MCS	Microsoft* Cluster Server
NIC	Network Interface Card
OS	Operating System
OPS	Oracle* Parallel Servers
PCI	Peripheral Component Interconnect
PHP	PCI Hot-plug
RAID	Redundant Array of Inexpensive Disks
SAN	Storage Area Network
SPA/B	Storage Processor. Referred to as SPA or SPB.
SP4/5	Service Pack. Referred to as SP4 or SP5.
Striping	Distributing data across several drives for redundancy
TB	Terabytes (approximately 1000 Billion=10^{12} bytes)
UPS	Un-interruptible Power Supply
VI	Virtual Interface Architecture
WWW	World Wide Web

References

Refer to the following Web sites for additional information:

Intel Corp:	http://www.intel.com
Fibre Channel*:	http://www.fibrechannel.com
CLARiiON*:	http://www.clariion.com
LSI LOGIC*:	http://www.lsil.com
Emulex*:	http://www.emulex.com
QLogic*:	http://www.qlc.com
DPT Corp.*:	http://www.dpt.com
Ancor*:	http://www.ancor.com
Cybex*:	http://www.cybex.com
OpenView*:	http://www.hp.com
Tivoli*:	http://www.tivoli.com
Unicenter*:	http://www.cai.com
Giganet*:	http://www.giganet.com
Fujitsu-HAL*:	http://www.fjst.com
IBM*	http://www.ibm.com
Intel VI	http://developer.intel.com/design/servers/vi/technology/specification.htm

Table of Contents

List of Figures

List of Tables

1. Abstract

This document describes Scalable Server Storage Solutions developed on four-way and eight-way Intel® Pentium® III Xeon™ processor based systems. Currently, this includes the AC450NX systems, based on the Intel® 450NX chipset, and the Intel® OCPRF100 systems, based on the Profusion* chipset. A brief description of the different server models is given in this paper, along with the actual solutions developed and equipment used.

Until recently, a scalable server storage solution on Intel® Architecture (IA) was very limited and did not provide the price/performance available on other platforms. With the advent of Fibre Channel (FC) storage and the availability of high performance four-way and eight-way Intel® Pentium® III Xeon™ processor based servers, it is now possible to configure and build enterprise level solutions with management functions that facilitate remote administration while providing a very flexible and highly scalable platform. Since 126 (125+1 host) devices can be attached to a single FC loop with 10 available PCI slots, the theoretical upper limit storage is 10x125x36G = 45,000 GB or 45 TB, using 36GB HDD units. Using 50G HDD units, which are recently becoming available, this limit increases to 62.5 TB attached to a **single server**. This kind of scalability was simply not possible using the conventional SCSI interface.

All of the server models developed here are based on Microsoft* Windows NT* 4.0 Enterprise Edition. Other operating systems (OS) can be adapted easily to these configurations as well. The primary storage interface used in these examples is Fibre Channel Arbitrated Loop (FC-AL). The advantages of FC-AL over the SCSI interface are too numerous and need not be listed here. If needed, please refer to the reference section for additional information on Fibre Channel.

2. Server Models

The simplest approach to a server solution is to provide a stand-alone server interfaced to the storage devices. The storage devices can be configured in either a “Just a Bunch of Disks (JBOD)” or a “Redundant Array of Inexpensive Disks (RAID)” solution.

At one end of the scale is JBOD. The advantage of JBOD is simple. Just attach the HDD units to the server, format/mount the HDD units, and begin storing data.

A RAID solution resides at the other end of the storage solution spectrum. The advantages of a RAID solution are “tunable” configurations for performance, automatic recovery if a HDD unit fails (no loss of data), and greater expandability. When compared to JBOD, a RAID solution is slightly more complex to set up. The basic RAID configurations are:

- RAID 0 Pure striping: no fail over, maximum use of drive size, highest performance.
- RAID 1 Drive mirroring: one-half of total drive size, complete redundancy.
- RAID 3 Striping with fixed parity: recoverable if drive fails, high bandwidth performance.
- RAID 5 Striping with floating parity: best configuration for I/O cost performance.
- RAID IO Disk striping with mirroring.

This document will focus on Fibre Channel as the interconnect between the server and the storage system. Fibre Channel is a serial interface that provides flexibility for scaling and is capable of attaching 126 devices per host adapter. Furthermore, FC hubs can be used in FC loops to share storage devices, thus allowing multiple servers to be clustered, providing greater availability to data.

Even though the solutions developed here are primarily based on RAID systems that are located outside the server, DPT has developed a FC PCI based RAID that is located internal to the server. Two storage solution suppliers have provided the storage systems used.

First, CLARiiON* has provided a FC5300 RAID system. The FC5300 is a storage solution that uses a dual controller, dual loop RAID solution with Fibre connection to the HDD units, and it is limited to a maximum of 30 HDD units. CLARiiON also provides the FC5000 JBOD system which includes a dual loop RAID solution with Fibre connection to the HDD units, and it is limited to a maximum of 120 drives.

Second, LSI Logic* has provided a MetaStor* RAID system. This system utilizes FC between the server and RAID controllers. Ultra 2 SCSI is then used between the RAID controllers and the HDD units.

The FC host adapters used in this development effort are QLogic* and Emulex*. The Emulex FC host adapter was utilized to attach to the CLARiiON storage system, and the Emulex FC host adapter and/or the Qlogic FC host adapter was utilized to attach to the LSI Logic storage system.

A factor that has increasing importance with server availability is the ability to Hot Replace PCI HBAs. Both the AC450NX and the OCPRF100 systems support this feature with the AC450NX having four hot-plug PCI slots and the OCPRF100 having 10 hot-plug PCI slots.** Failed HBAs can be powered-down, removed, replaced, and the replacement powered-up without interrupting other server functions.

2.1 Single Server

The most common storage solution is based on a single server attached to a storage subsystem. The storage can be either JBOD- or RAID-based. The availability of data is the determining factor when choosing a solution. JBOD allows for the storage of data at the lowest cost but without recoverability if a HDD unit should fail. RAID solutions increase the cost but provide the ability to recover data if problems arise.

With all of the single server configurations, common building blocks are shared. The building blocks consist of Microsoft Windows NT 4.0 Server installed on the system along with Network options (TCP/IP, SNMP, etc.), storage HBA drivers, and if needed, RAID utilities. There is a network interface card (NIC) which connects the server to the rest of the network. HP* OpenView*, Tivoli*, or Computer Associate* (CA) UniCenter* software is installed on a client workstation to allow system management, fault isolation, and security-related functions.

The typical single server/single storage model is as follows:

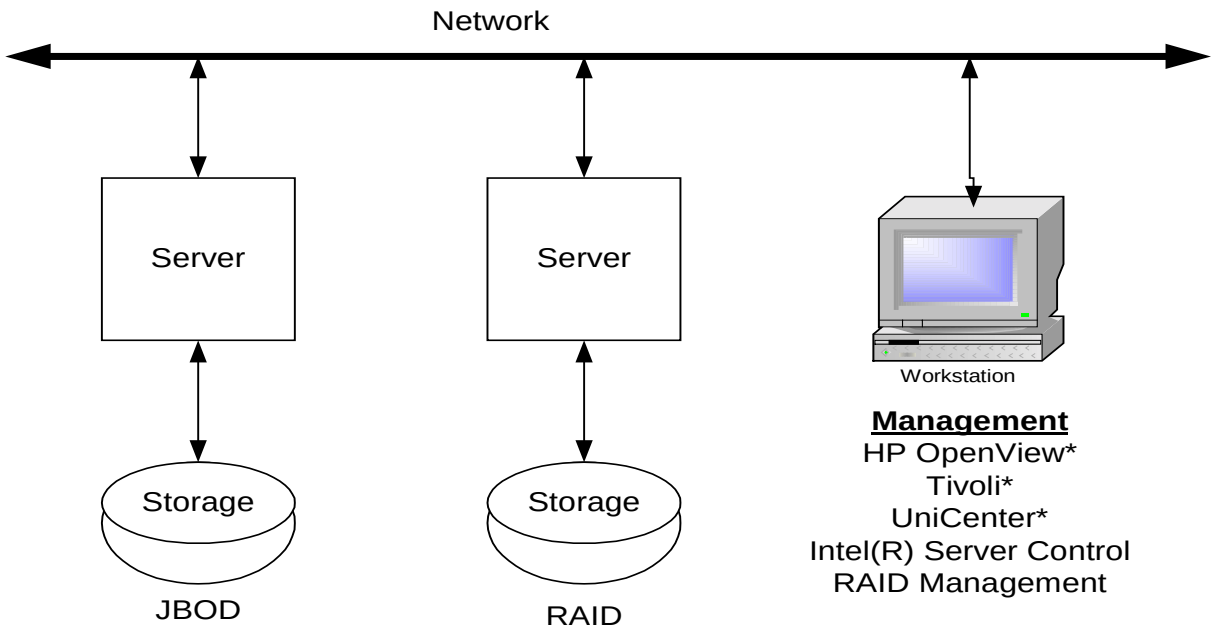

Figure 1: Single Server/Dedicated Storage Model

The four cases that were used for this model are detailed below.

2.1.1 JBOD System

JBOD is the simplest storage configuration. JBOD refers to “just a bunch of disks” and it is HDD units attached to the FC loop. The number of HDD units attached is limited the number of HDD unit letters usable to Windows NT 4.0. If you need more HDD units than the quantity of letters in the alphabet, a RAID is the next step in your storage solution.

After the storage HBA is installed, its driver is loaded, and the HDD units are attached by executing Disk Administrator. Use the tools in disk administrator to format and assign letters to all the HDD units. Reboot the system and start storing data.

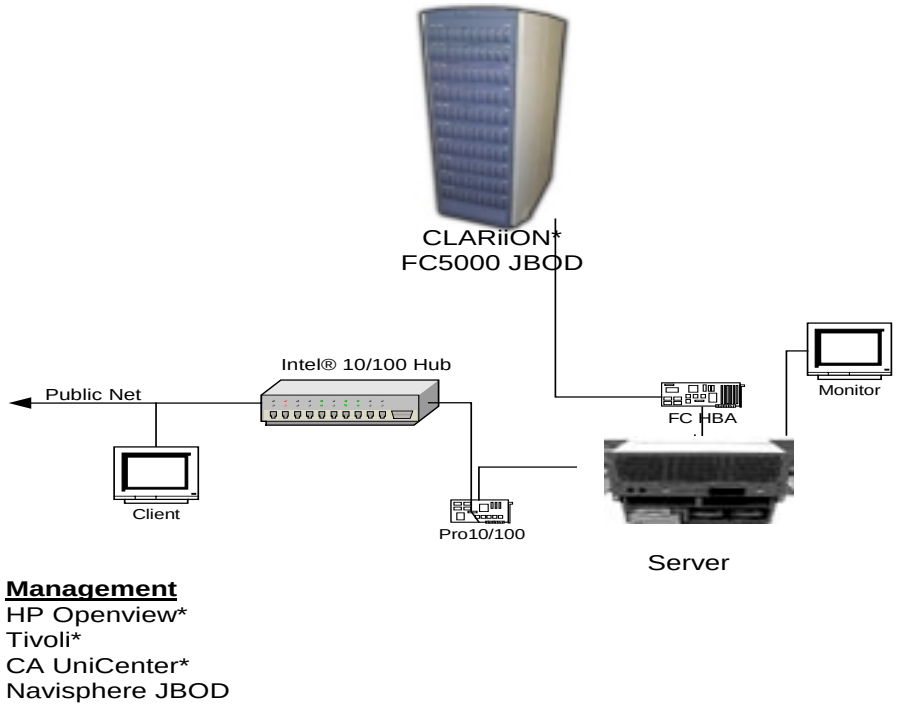

Figure 2: Single Server JBOD System

The CLARiiON JBOD solution uses a Qlogic QLA2100/66 or an Emulex LP8000 FC PCI Adapter. The adapter is connected to a CLARiiON Disk Array Enclosure (DAE). The two purposes of the NIC card are to share stored data and provide remote administration by a workstation.

System Summary

The following table summarizes the various components used to assemble the system. Note that the client systems are not included in this list.

Table 1: Single Server JBOD System Summary

Item	Model	Part #	Qty	Notes
Server: Intel Intel® AC450NX System, four 550-MHz Pentium® III Xeon™ processors, 1GB RAM **OR** Intel® OCPRF100 System, eight 550-MHz Pentium® III Xeon™ processors, 1GB RAM	AC450NX OCPRF100		1	System uses onboard SCSI drives to boot to Windows NT* 4.0 Enterprise Edition. Hot-plug PCI slots for “like for like“ replacement.
Storage: CLARiiON* DAE Disk: CLARiiON Japanese and Korean PN# DAE Disk: CLARiiON	 DAE ST39102FC DAE ST39102FC	 C5051R-A C8810FLG-A C5051R-D C8810FLG-A	 1 10 1 10	
FC HBA: Emulex* FW 2.81 Qlogic* BIOS 1.35	 LP8000 Qla2100/66	 LP8000-T1 Qla2100/66-BK	1	
FC cable: Amp Emulex style Qlogic style	 5M DB9-DB9 3M HSSDC-DB9	 621771-6 636246-2	1	See FC specifications for more info.
Ethernet NIC: Intel	Intel® EtherExpress ™ PRO/100+	PILA8470B	1	Ethernet 10/100Mbps
Ethernet HUB: Intel	Intel® Express 140T	EE140TX24 US	1	Running at 100Mbs.
Ethernet CAT5 Cable	5 to 15 feet		1	Longer lengths may be needed
Operating system	Windows NT 4.0 SP5			Microsoft*
Enterprise management: Choose one-	HP OpenView* Tivoli* UniCenter*		1	Hewlett-Packard* Tivoli CAI*

2.1.2 DPT* System

DPT* has provided a Fibre Channel PCI based RAID controller. DPT* has provided this storage solution by introducing the model 3755FC/ SX4055F dual channel RAID controller. This product will enable the user to configure all of the HDD units for data recovery.

Install the 3755FC/ SX4055F, the driver, and “DPT Storage Manager V2.12” into the server. Attach each LCC in the DAE on the 3755FC/ SX4055F controller to the HBA. Configure the RAID using Storage Manager, choosing either auto-configure or manual-configure.

Note: See References for more specific information.

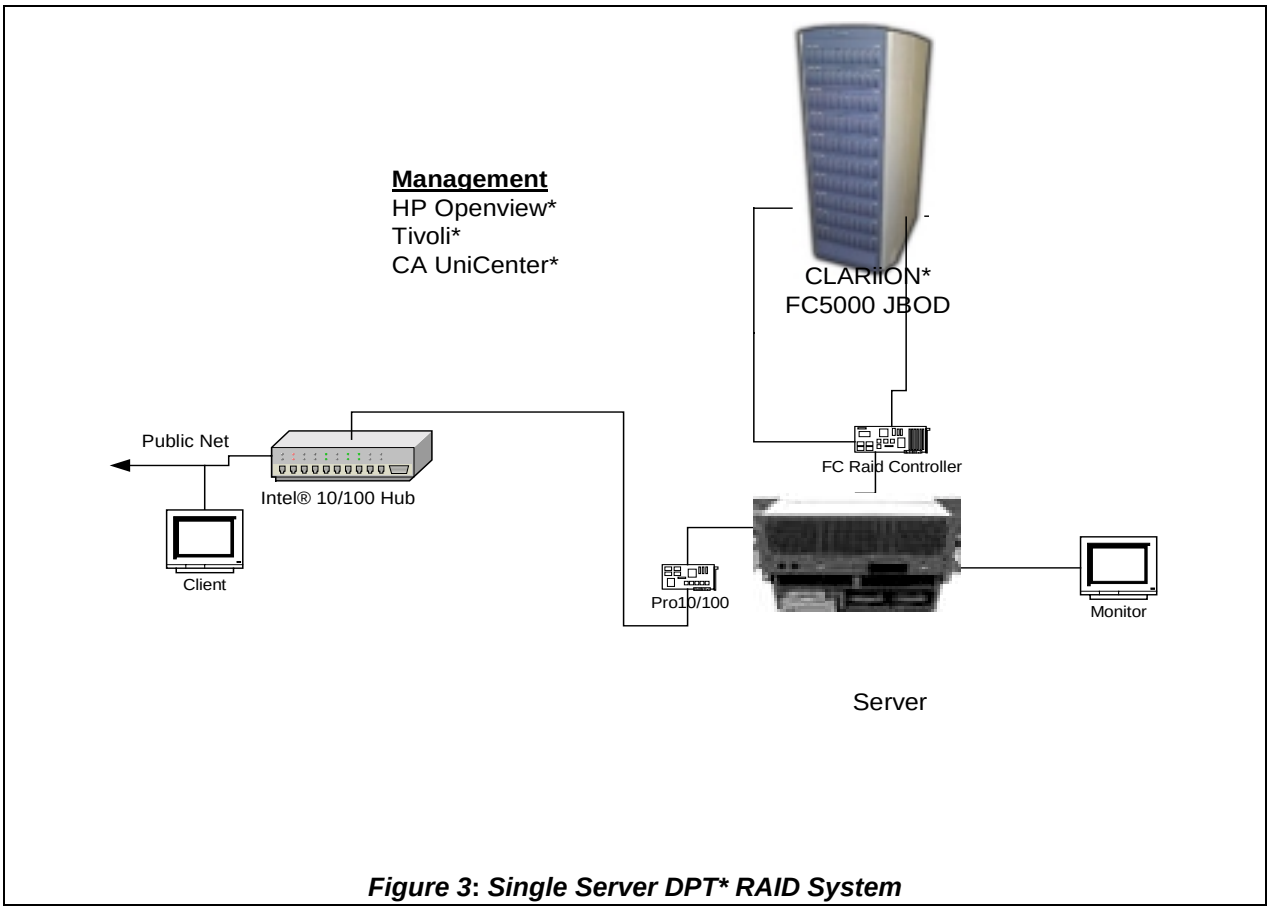

Figure 3*: *Single Server DPT* RAID System

The DPT FC RAID Controller has a Graphical User Interface (GUI) which controls the card. This enables fast setup of RAID configurations and management of the HDD units. The controller is connected to a DAE through both LCCs.

The following table summarizes the various components used to assemble the system. Note that the client systems are not included in this list.

Table 2: Single Server DPT* System Summary

Item	Model	Part #	Qty	Notes
Server: Intel Intel® AC450NX System, four 550-MHz Pentium® III Xeon™ processors, 1GB RAM **OR** Intel® OCPRF100 System, eight 550-MHz Pentium® III Xeon™ processors, 1GB RAM	AC450NX OCPRF100		1	System uses onboard SCSI drives to boot to Windows NT* 4.0 Enterprise Edition. Hot-plug PCI slots for “like for like“ replacement.

Item	Model	Part #	Qty	Notes
Storage: CLARiiON*				
DAE	DAE	C5051R-A	1	
Disk: CLARiiON	ST39102FC	C8810FLG-A	10	
Japanese and Korean PN#				
DAE	DAE	C5051R-D	1	
Disk: CLARiiON	ST39102FC	C8810FLG-A	10	
FC RAID HBA: DPT*				
Base RAID HBA	PM3755F	PM3755F	1	
Extension Module	SX4055F	SX4055F	1	
FC cable: Amp	3M HSSDC-DB9	636246-2	2	See FC specifications for more info.
Ethernet NIC: Intel	Intel® EtherExpress ™ PRO/100+	PILA8470B	1	Ethernet 10/100Mbps
Ethernet HUB: Intel	Intel® Express 140T	EE140TX24US	1	Running at 100Mbs.
Ethernet CAT5 Cable	5 to 15 feet		1	Longer lengths may be needed
Operating system	Windows NT 4.0 SP5			Microsoft*
Enterprise management:	HP OpenView*		1	Hewlett-Packard*
Choose one-	Tivoli*			Tivoli
	UniCenter*			CAI*

2.1.3 CLARiiON* System

The CLARiiON FC5300 storage system features a dual loop, dual controller FC host interface. The dual active controllers feature mirrored write cache for reliability and high performance. The redundant components greatly enhance the availability of the system by providing alternate paths from the storage to the host system. All components, including fans, power supplies, and HDD units, can be hot-plugged and hot-replaced. The CLARiiON FC5300 can also support up to thirty HDD units installed in three DAEs. Combining these two features results in a highly available and scalable storage enclosure. Also, the FC5300 can be scaled into a FC5700 system with up to 120 drives. With the addition of dual FC HBAs located in the hot-plug PCI slots, each HBA connected to a separate LCC, a highly available storage system has been generated.

For enterprise-level management, HP OpenView, Tivoli or CA UniCenter is installed on the client system. Intel® Server Control software provides component level fault isolation of the server and the CLARiiON Navisphere Manager provides the RAID management functions from the workstation. Navisphere also provides status monitoring and fault isolation for the RAID hardware modules such as controllers, fans, FC link cards, power supplies, and HDD units.

To configure a system, install the FC HBAs into the PCI slots; install the drivers and Navisphere software, and connect to the CLARiiON FC5300. If using the AC450NX, install the FC HBAs into the 64-bit slots for high performance. If using the OCPRF100 system, install each FC HBA into separate PCI segments (segment C and segment D) for high bandwidth and additional availability.

Note: Reference the CLARiiON documentation on RAID configuration.

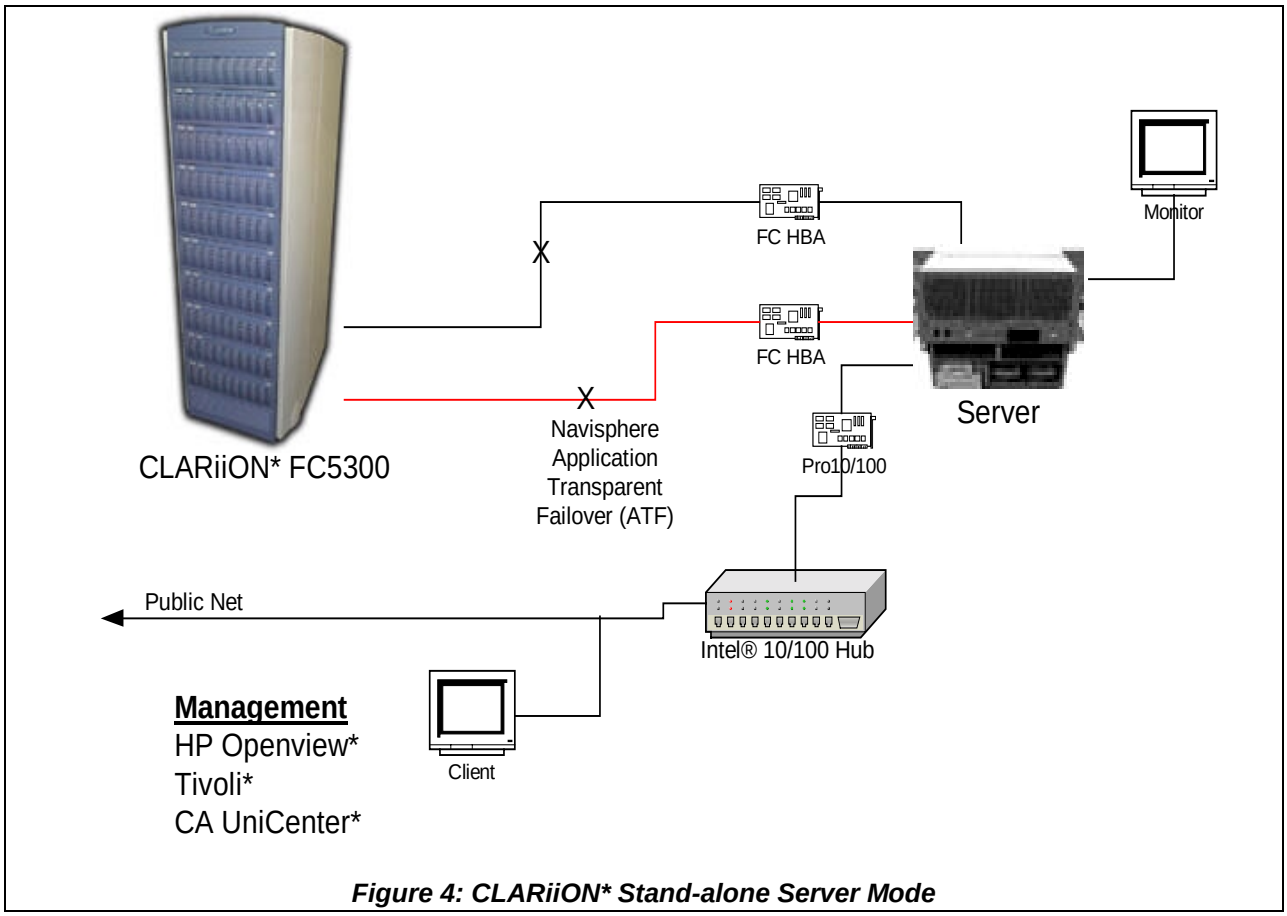

Figure 4: CLARiiON* Stand-alone Server Mode

The current PHP solution supports “like for like” replacement which means a PCI card can be replaced with an identical version and continue using the same drivers.

System Summary

The following table summarizes the various components used to assemble the system. Note that the client systems are not included in this list.

Table 3: CLARiiON* Stand-alone System Summary

Item	Model	Part #	Qty	Notes
Server: Intel Intel® AC450NX System, four 550-MHz Pentium® III Xeon™ processors, 1GB RAM **OR** Intel® OCPRF100 System, eight 550-MHz Pentium® III Xeon™ processors, 1GB RAM	AC450NX OCPRF100		1	System uses onboard SCSI drives to boot to Windows NT* 4.0 Enterprise Edition. Hot-plug PCI slots for “like for like“ replacement.
Storage: CLARiiON* FC5300 Navisphere Application Transparent Failover		C259BBG1CD	1	Order qty 1 per server.

Item	Model	Part #	Qty	Notes
RAID LCC w/ DAE	5300	C5301R-A	1	SP's w/ FW 5.21.02
DAE	DAE	C5051R-A	2	
Cabinet	C7661G-F7B	C7661G-F7B	1	
Disk: CLARiiON Firmware 3528	ST39102FC	C8810FLG-A	30	
Dual Stand-by Power Supply	SPS for 5300	C7214G-A	1	
Cache DIMM	128MB Cache	C72128G-A	2	
Attach Kit w/ Emulex* LP8000 FC HBA and software	NT Attach Kit	S5K53M-N10	2	
Japanese and Korean PN#				
RAID LCC w/ DAE	5300	C5301R-D	1	
DAE	DAE	C5051R-D	2	
Cabinet	C7661G-F7A	C7661G-F7A	1	
Disk: CLARiiON Firmware 3528	ST39102FC	C8810FLG-A	30	
Dual Stand-by Power Supply	SPS for 5300	C7214G-D	1	
Cache DIMM	128MB Cache	C72128G-A	2	
Attach Kit w/ Emulex LP8000 FC HBA and software	NT Attach Kit	S5K53M-N10	2	
FC cable	5M DB9-DB9		2	Longer lengths may be needed. See FC specifications for more info.
Ethernet NIC: Intel	Intel® EtherExpress™ PRO/100+	PILA8470B	1	Ethernet 10/100Mbps
Ethernet HUB: Intel	Intel® Express 140T	EE140TX24US	1	Running at 100Mbs.
Ethernet CAT5 Cable	5 to 15 feet		1	Longer lengths may be needed
Operating system	Windows NT 4.0 SP5			Microsoft*
RAID management	Navisphere*			CLARiiON V4.0
Enterprise management: Choose one-	HP OpenView* Tivoli* UniCenter*		1	Hewlett-Packard* Tivoli CAI*

2.1.4 LSI Logic* System

The LSI Logic MetaStor storage system utilizes a dual loop, dual controller solution. The LSI Logic MetaStor uses a Fibre Channel interface between the server and controller, and an LVD SCSI interface between the controller and the HDD units. The LSI Logic MetaStor dual loop /dual controller system results in a highly available and very scalable storage solution. The host interface is FC and both Emulex and Qlogic HBAs are compatible with this storage system. SYMplicity* Storage Manager runs on a client workstation and provides the RAID management features.

The LSI Logic MetaStor 4766 controller has an internal 10/100Mbs Ethernet Interface for remote system management. The 4766 controller also has a serial connector for similar use with a null modem connection.

To install an LSI Logic MetaStor storage system, install the FC HBAs into the PCI slots, connect to the controllers, install the drivers, and install SYMplicity RAID software. When using the AC450NX, install the FC HBAs into the 64-bit slots for high performance. When using the

OCPRF100 system, install each FC HBA into separate PCI segments (segment C and segment D) for high bandwidth and additional availability.

Note: The LSI Logic documentation covers how to set up the HDD units for different RAID configurations.

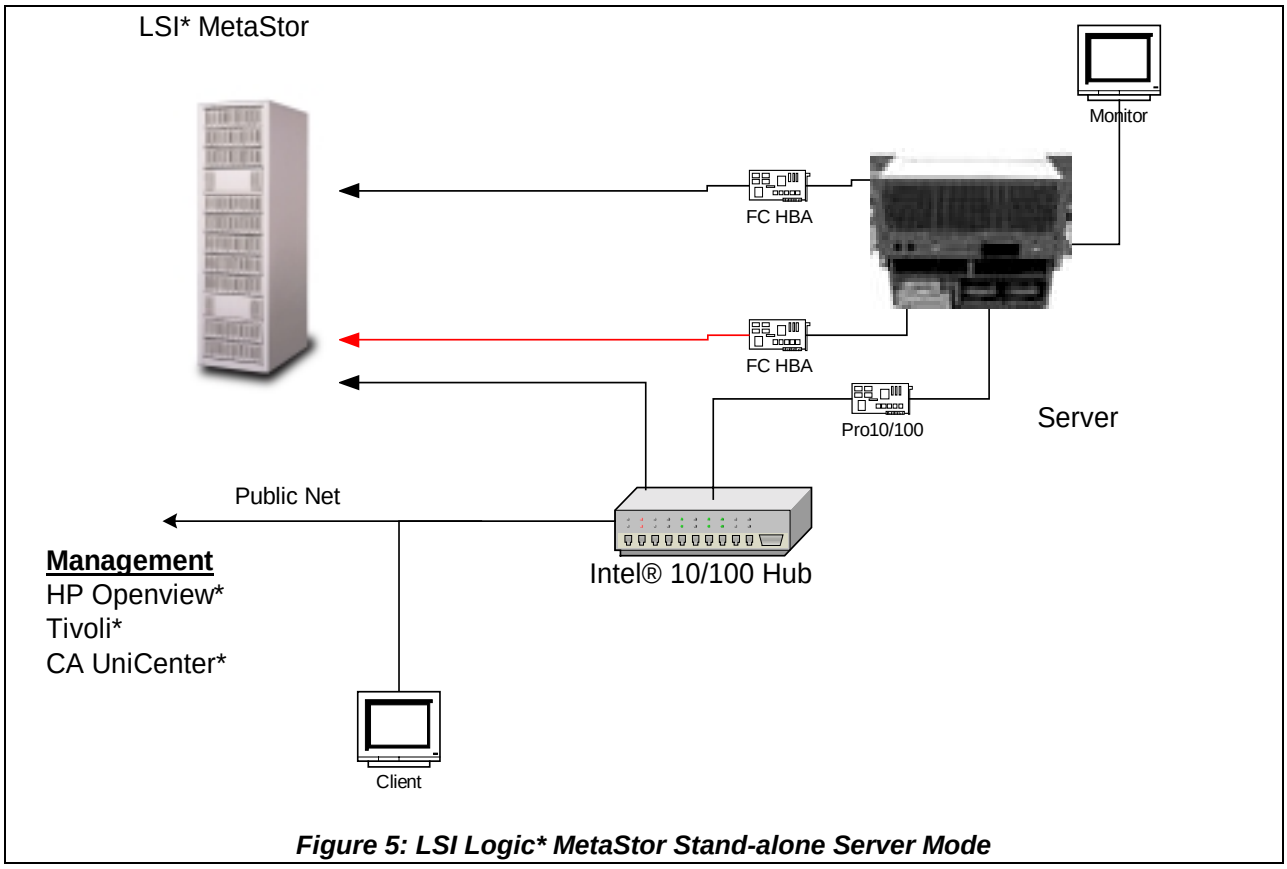

Figure 5: LSI Logic* MetaStor Stand-alone Server Mode

System Summary

The following table summarizes the various components used to assemble the system. Note that the client systems are not included in this list.

Table 4: LSI Logic* MetaStor Stand-alone System Summary

Item	Model	Part #	Qty	Notes
Server: Intel Intel® AC450NX System, four 550-MHz Pentium® III Xeon™ processors, 1GB RAM **OR** Intel® OCPRF100 System, eight 550-MHz Pentium® III Xeon™ processors, 1GB RAM	AC450NX OCPRF100		1	System uses onboard SCSI drives to boot to Windows NT* 4.0 Enterprise Edition. Hot-plug PCI slots for “like for like“ replacement.
Storage: LSI Logic*				

Item	Model	Part #	Qty	Notes
72 inch Cabinet	Rack Mount	6256-4000-6600	1	
Rack Mount Rail Sets	DM/CM	6256-F958-0000	4	
Command Module	Rack for 4766	1000-3101-6600	1	
Controller	4766 256MB	1000-F120-0000	2	
Drive Module	Rack Mount	2000-2101-6600	3	
SCSI Connect CRU	LVD-SE	2000-F038-000	3	
Drives, 9GB U2	10K	2000-F210-0000	30	
Controller Code3.X	Software	1000-F951-0000	1	
Fiber Optic cable	MM	006-1086417	2	
FC HBA: Emulex* FW 2.81 Qlogic* BIOS 1.35	 LP8000 Qla2100F/66	 LP8000-N1 Qla2100F/66-BK	2	
Ethernet NIC: Intel	Intel® EtherExpress™ PRO/100+	PILA8470B	1	Ethernet 10/100Mbps
Ethernet HUB: Intel	Intel® Express 140T	EE140TX24US	1	Running at 100Mbs.
Ethernet CAT5 Cable	5 to 15 feet		2	Longer lengths may be needed
Operating system	Windows NT 4.0 SP5			Microsoft*
Enterprise management: Choose one-	HP OpenView* Tivoli* UniCenter*		1	Hewlett-Packard* Tivoli CAI*

2.2 Microsoft* Cluster Servers Model

The clustered server model has the advantage of a “fail-over” mechanism to prevent loss of server functions when one server malfunctions or crashes. One clustered server solution that is available for Windows NT systems is provided in the Windows NT 4.0 Enterprise Edition. The solution developed was Microsoft Cluster Server**,** and this has been used on both the CLARiiON FC5300 and the LSI LOGIC MetaStor systems.

To insure greater availability of a server/storage system, redundant pieces of hardware are added so that if one device fails, another is there for fail-over. In a single server model, if a drive fails, the RAID system will recover the data. If a RAID controller or a FC HBA fails, another is there for fail-over. The next step is to add a second server, and using it for fail-over.

The following diagram shows the model of this configuration.

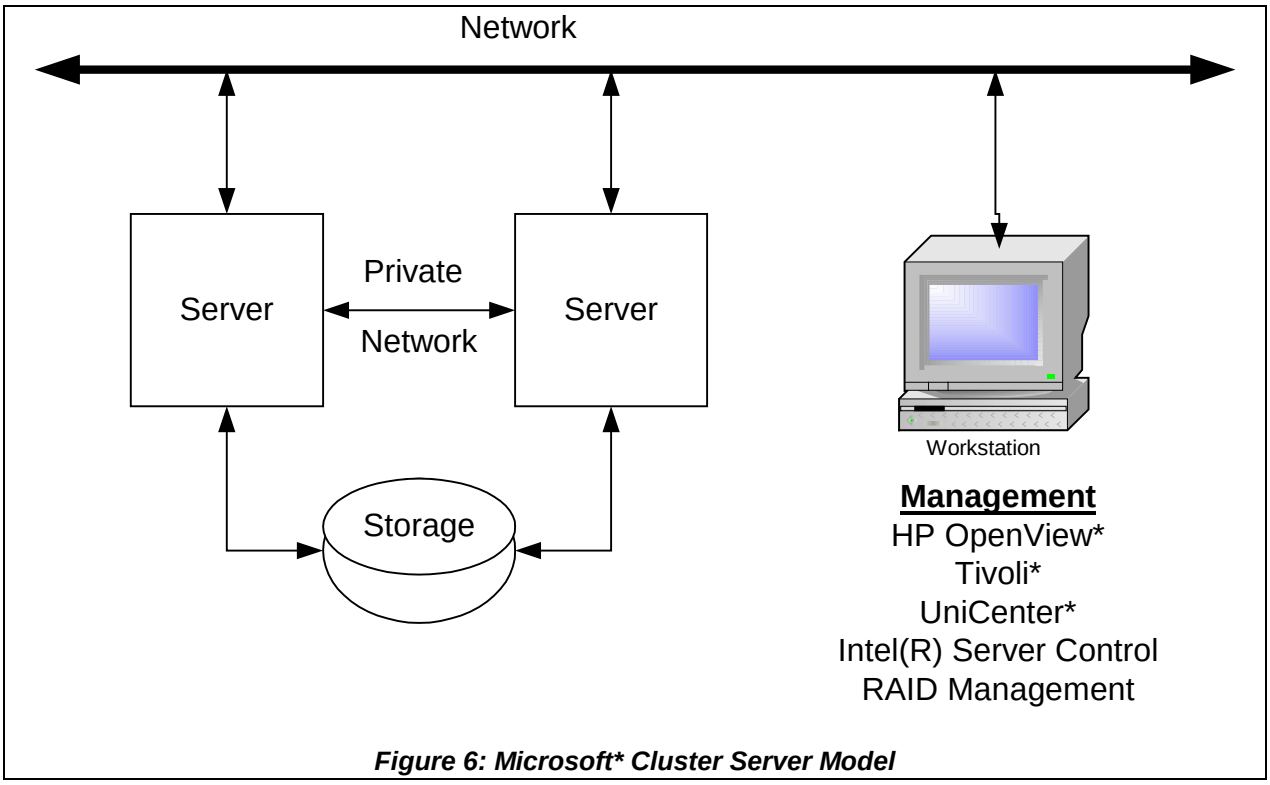

Figure 6: Microsoft* Cluster Server Model

In this model, the private Ethernet network passes “keep alive” information between the two servers. The private network requires two Network Interface Cards (NICs) per server, each NIC is connected to a separate hub. An additional NIC is installed in each system for communication with the public network. A management workstation will manage both servers in the cluster from the public net using HP OpenView, Tivoli, or Unicenter.

Complete storage redundancy is achieved by using dual FC HBAs installed in the PHP slots of each server. If a failure occurs, replacement of the failed HBA is seamless requiring neither server to be taken off-line.

2.2.1 CLARiiON*

The clustered configuration used for the CLARiiON FC5300 uses the single server model as its building blocks. This configuration consists of two single server FC5300 systems, with the additional MCS software and a dual private network. Connect the two FC HBAs in each system to separate FC5300 SPs. To be more specific, connect Server1 HBA1 to FC5300 SPA and Server1 HBA2 to FC5300 SPB. The connections should be the same for server2. The CLARiiON FC5300 system contains a mini-hub that is built into the RAID SP. No external hub is required when attaching up to four HBAs.

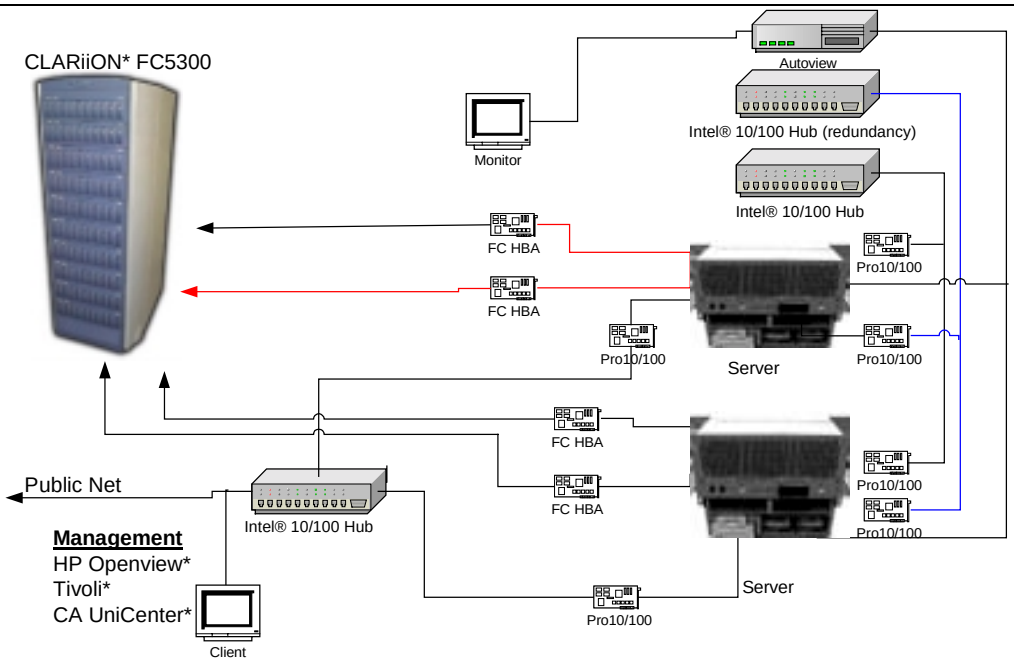

Figure 7: CLARiiON* FC5300, Microsoft* Cluster Server Cluster

Windows NT 4.0 Enterprise Edition provides all the components needed to run a clustered server solution. In this mode, both servers provide a pool of services and either server can take over the host functions should one server fail to continue reliable operation.

The following table summarizes the various hardware and software components used to assemble this configuration.

Table 5: CLARiiON* Microsoft* Cluster Server System Summary

Item	Model	Part #	Qty	Notes
Server: Intel Intel® AC450NX System, four 550-MHz Pentium® III Xeon™ processors, 1GB RAM **OR** Intel® OCPRF100 System, eight 550-MHz Pentium® III Xeon™ processors, 1GB RAM	AC450NX OCPRF100		2	System uses onboard SCSI drives to boot to Windows NT* 4.0 Enterprise Edition. Hot-plug PCI slots for “like for like“ replacement.
Monitor Switch: Cybex* Cybex Cables	AutoView Commander 8 ft monitor, mouse, kb	AV-8B CUFC-8	1 2	See Cybex for more info.
Storage: CLARiiON* FC5300 Navisphere Application Transparent Failover RAID LCC w/ DAE DAE	 5300 DAE	 C259BBG1CD C5301R-A C5051R-A	 1 1 2	 Order qty 1 per server. SP’s w/ FW 5.21.02

Item	Model	Part #	Qty	Notes
Cabinet	C7661G-F7B	C7661G-F7B	1	
Disk: CLARiiON Firmware 3528	ST39102FC	C8810FLG-A	30	
Dual Stand-by Power Supply	SPS for 5300	C7214G-A	1	
Cache DIMM	128MB Cache	C72128G-A	2	
Attach Kit w/ Emulex* LP8000 FC HBA and software	NT Attach Kit	S5K53M-N10	4	
Japanese and Korean PN#				
RAID LCC w/ DAE	5300	C5301R-D	1	
DAE	DAE	C5051R-D	2	
Cabinet	C7661G-F7A	C7661G-F7A	1	
Disk: CLARiiON Firmware 3528	ST39102FC	C8810FLG-A	30	
Dual Stand-by Power Supply	SPS for 5300	C7214G-D	1	
Cache DIMM	128MB Cache	C72128G-A	2	
Attach Kit w/ Emulex LP8000 FC HBA and software	NT Attach Kit	S5K53M-N10	4	
FC cable	5M DB9-DB9		4	Longer lengths may be needed. See FC specifications for more info.
Ethernet NIC: Intel	Intel® EtherExpress™ PRO/100+	PILA8470B	1	Ethernet 10/100Mbps
Ethernet HUB: Intel	Intel® Express 140T	EE140TX24US	1	Running at 100Mbs.
Ethernet CAT5 Cable	5 to 15 feet		6	Longer lengths may be needed
Operating system	Windows NT 4.0 SP5			Microsoft*
Microsoft Cluster Server	MCS			Microsoft
RAID management	Navisphere*			CLARiiON V4.0
Enterprise management: Choose one-	HP OpenView* Tivoli* UniCenter*		1	Hewlett-Packard* Tivoli CAI*

2.2.2 LSI Logic*

The clustered configuration used for LSI Logic* uses the single server model as its building blocks. The addition of the redundant, private network and the MCS software completes the server cluster. Additionally, a FC hub has also been added to each side due to the single port design of each RAID controller. Connect the two FC HBAs in each system to separate hubs. To be more specific; connect Server1 HBA1 to hub1 and Server1 HBA2 to hub2, and likewise for server2. Utilizing this completely redundant system, a highly available storage solution will be produced.

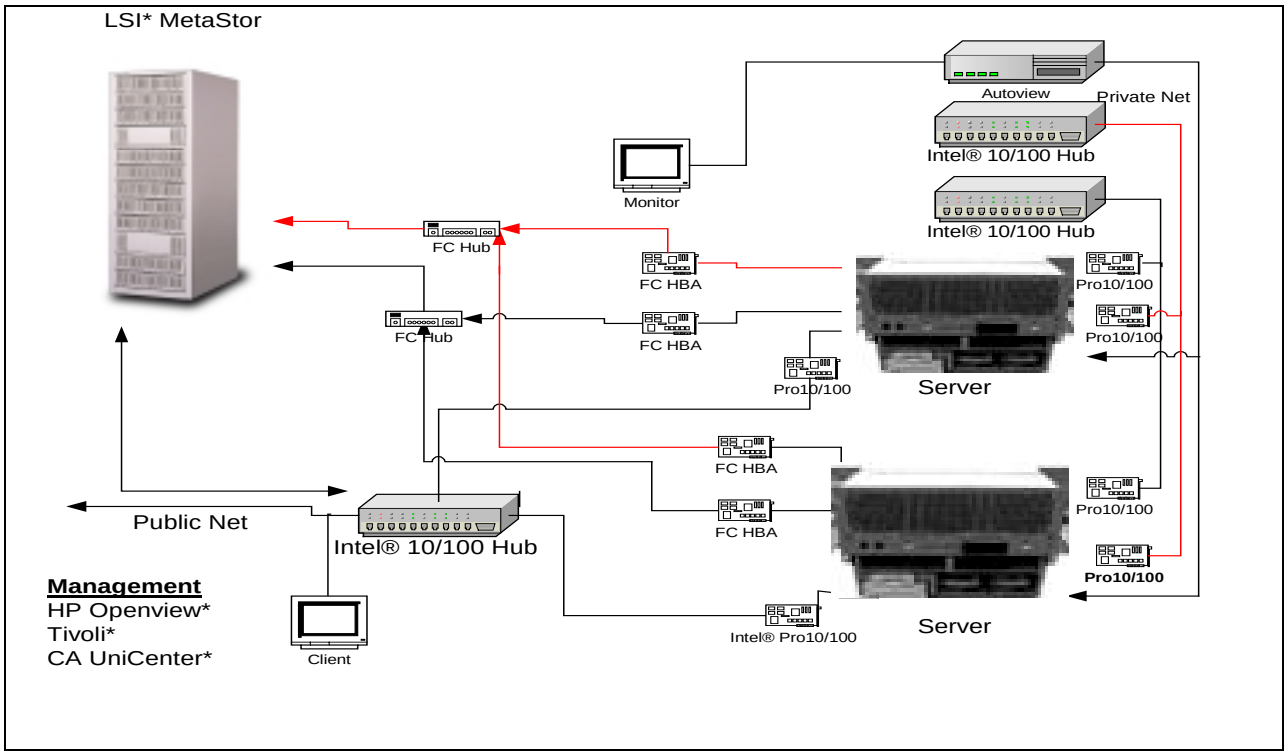

Figure 8: LSI Logic* MetaStor, Microsoft* Cluster Server Cluster

Windows NT 4.0 Enterprise Server Edition provides the software components required to run clustered servers. In this mode, both servers provide a pool of services and either server can take over the host functions should the other server fail to continue reliable operation.

The following is a summary of the configuration.

Table 6: LSI Logic* Microsoft* Cluster Server System Summary

Item	Model	Part #	Qty	Notes
Server: Intel Intel® AC450NX System, four 550-MHz Pentium® III Xeon™ processors, 1GB RAM **OR** Intel® OCPRF100 System, eight 550-MHz Pentium® III Xeon™ processors, 1GB RAM	AC450NX OCPRF100		2	System uses onboard SCSI drives to boot to Windows NT* 4.0 Enterprise Edition. Hot-plug PCI slots for “like for like“ replacement.
Storage: LSI Logic*				
72 inch Cabinet	Rack Mount	6256-4000-6600	1	
Rack Mount Rail Sets	DM/CM	6256-F958-0000	4	
Command Module	Rack for 4766	1000-3101-6600	1	
Controller	4766 256MB	1000-F120-0000	2	

Item	Model	Part #	Qty	Notes
Drive Module	Rack Mount	2000-2101-6600	3	
SCSI Connect CRU	LVD-SE	2000-F038-000	3	
Drives, 9GB U2	10K	2000-F210-0000	30	
Controller Code3.X	Software	1000-F951-0000	1	
Fiber Optic cable	MM	006-1086417	2	
FC HBA: Emulex FW 2.81 Qlogic BIOS 1.35	 LP8000 Qla2100F/66	 LP8000-N1 Qla2100F/66-BK	4	
FC Hub: Emulex*	LH5000	LH5000	2	Managed through 10Mb Ethernet
GBICs for Hub	Optical	LGB 100-N1	6	
Fiber Optic Cable: QFC	MM 3M	Q2xM2203	4	
FC cable: Amp	5M DB9-DB9	621771-6	2	Used between MetaStor* and FC Hubs
Ethernet NIC: Intel	Intel® EtherExpress™ PRO/100+	PILA8470B	1	Ethernet 10/100Mbps
Ethernet HUB: Intel	Intel® Express 140T	EE140TX24US	1	Running at 100Mbs.
Ethernet CAT5 Cable	5 to 15 feet		6	Longer lengths may be needed
Operating system	Windows NT 4.0 SP5			Microsoft*
Microsoft Cluster Server	MCS			Microsoft
Enterprise management: Choose one-	HP OpenView* Tivoli* UniCenter*		1	Hewlett-Packard* Tivoli CAI*

2.3 Oracle* Parallel Server Model

The Oracle* Parallel Server (OPS) model is based on the Oracle8I* and Parallel Server software combined with the power of the cluster to provide high availability, scalability and single system manageability. The Oracle caching technology that reduces HDD unit accessing and exploits the bandwidth with low latency interconnects and provides increased scalability has enhanced these three areas. These enhancements are included in OPS mission-critical online transaction processing (OLTP).

High availability protects the user and application from hardware and software down time. Virtual Interface Architecture (VIA) is a new high speed interconnect that increases the speed of the transaction by allowing all the servers to pass information as required to complete the request.

Cluster scalability allows more users to request more processing throughput. This is accomplished while expanding the requirement for more CPU, memory and HDD unit capacity. The more servers on the cluster the greater number of transactions processed.

Single system management brings simplicity to the managing of the cluster by viewing it as a single system. The administrator is not confused as to where the data resides.

The cluster requires a high speed interconnect (VI) that connects with the servers. This allows control and smooth flow of the transactions.

NICs connect the server to the Administrator for management of the cluster and connection to the clients.

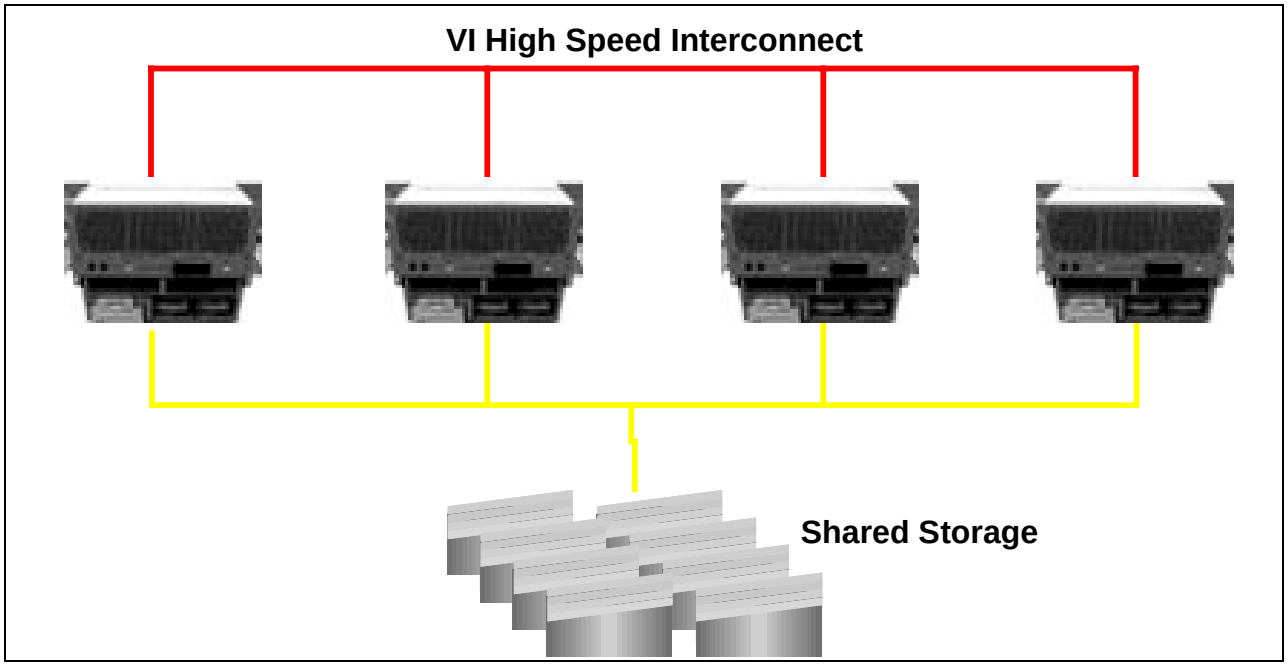

Figure 9: Oracle* Parallel Server Cluster Model

2.3.1 CLARiiON*

With four enterprise servers running OPS connected to the CLARiiON FC5300, a major increase in scalability is possible. Navisphere was used to configure the HDD units into logical units required by Oracle. Partition the logical units using Windows NT 4.0's disk administrator. Execute the utility programs located in Oracle, and OPS latches the HDD units.

Note: Reference OPS documentation for additional information.

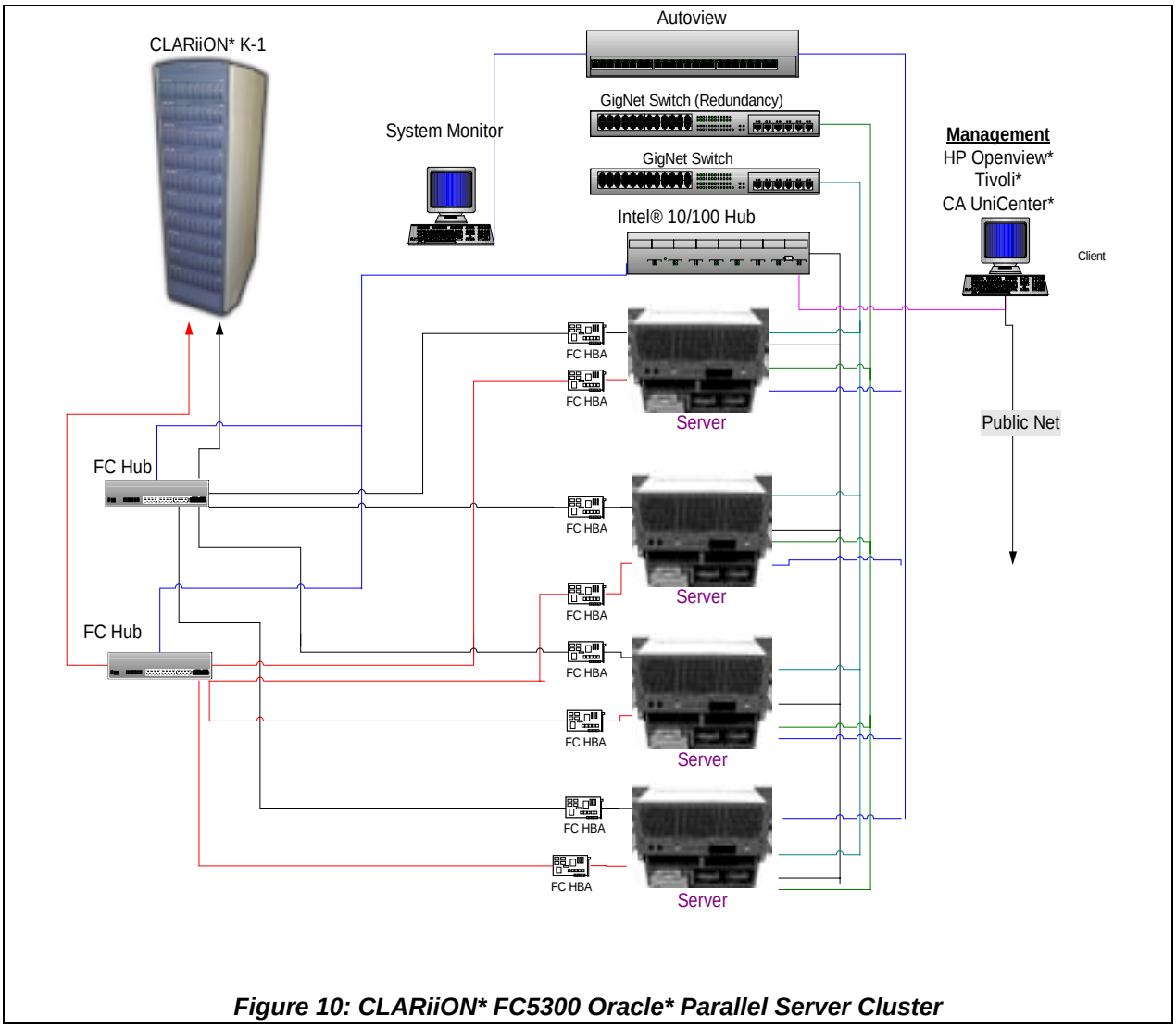

Figure 10: CLARiiON* FC5300 Oracle* Parallel Server Cluster

System Summary

The following table shows the necessary equipment to assemble the configuration.

Table 7: CLARiiON* Oracle* Parallel Server System Summary

Item	Model	Part #	Qty.	Notes
Server: Intel Intel® AC450NX System, four 550-MHz Pentium® III Xeon™ processors, 1GB RAM **OR** Intel® OCPRF100 System, eight 550-MHz Pentium® III Xeon™ processors, 1GB RAM	AC450NX OCPRF100		4	System uses onboard SCSI drives to boot to Windows NT* 4.0 Enterprise Edition. Hot-plug PCI slots for “like for like“ replacement.
Monitor Switch: Cybex*	AutoView Commander	AV-8B	1	See Cybex for more info.

Item	Model	Part #	Qty.	Notes
Cybex Cables	8 ft monitor, mouse, kb	CUFC-8	4	
Storage: CLARiiON* FC5300				
Navisphere Application Transparent Failover		C259BBG1CD	1	Order qty 1 per server.
RAID LCC w/ DAE	5300	C5301R-A	1	SP's w/ FW 5.21.02
DAE	DAE	C5051R-A	2	
Cabinet	C7661G-F7B	C7661G-F7B	1	
Disk: CLARiiON Firmware 3528	ST39102FC	C8810FLG-A	30	
Dual Stand-by Power Supply	SPS for 5300	C7214G-A	1	
Cache DIMM	128MB Cache	C72128G-A	2	
Attach Kit w/ Emulex* LP8000 FC HBA and software	NT Attach Kit	S5K53M-N10	8	
Japanese and Korean PN#				
RAID LCC w/ DAE	5300	C5301R-D	1	
DAE	DAE	C5051R-D	2	
Cabinet	C7661G-F7A	C7661G-F7A	1	
Disk: CLARiiON Firmware 3528	ST39102FC	C8810FLG-A	30	
Dual Stand-by Power Supply	SPS for 5300	C7214G-D	1	
Cache DIMM	128MB Cache	C72128G-A	2	
Attach Kit w/ Emulex LP8000 FC HBA and software	NT Attach Kit	S5K53M-N10		
FC Hub: Emulex	LH5000	LH5000	2	Managed through 10Mb Ethernet
GBICs for Hub	Cu DB9	LGB 100-T1	8	2ea DB9 connectors included w/ hub
FC cable: Copper	5M DB9-DB9		10	Longer lengths may be needed. See FC specifications for more info.
VI HBA: Giganet	cLAN1000	cLAN1000	8	2 HBA per system
VI Switch: Giganet	cLAN5000	cLAN5000	2	
VI cables	2M HSSDC-HSSDC	cLAN-A0211	8	
Ethernet NIC: Intel	Intel® EtherExpress™ PRO/100+	PILA8470B	1	Ethernet 10/100Mbps
Ethernet HUB: Intel	Intel® Express 140T	EE140TX24US	1	Running at 100Mbs.
Ethernet CAT5 Cable	5 to 15 feet		7	Longer lengths may be needed
Operating system	Windows NT 4.0 SP5			Microsoft*
Data Base: Oracle*	Oracle Parallel Server 8.1			Contact Oracle for specific application needs
RAID management	Navisphere*			CLARiiON V4.0
Enterprise management:	HP OpenView*		1	Hewlett-Packard*
Choose one-	Tivoli*			Tivoli
	UniCenter*			CAI*

2.3.2 LSI Logic*

With four enterprise servers running OPS connected to the LSI Logic MetaStor, a major increase in scalability is possible. The HDDs are configured using MetaStor's RAID utility into

the logical units required by Oracle. Next, partition the logical units using Windows NT 4.0's disk administrator. Execute the utility programs located in Oracle, and OPS latches the HDD units.

Note: Reference OPS documentation for additional information.

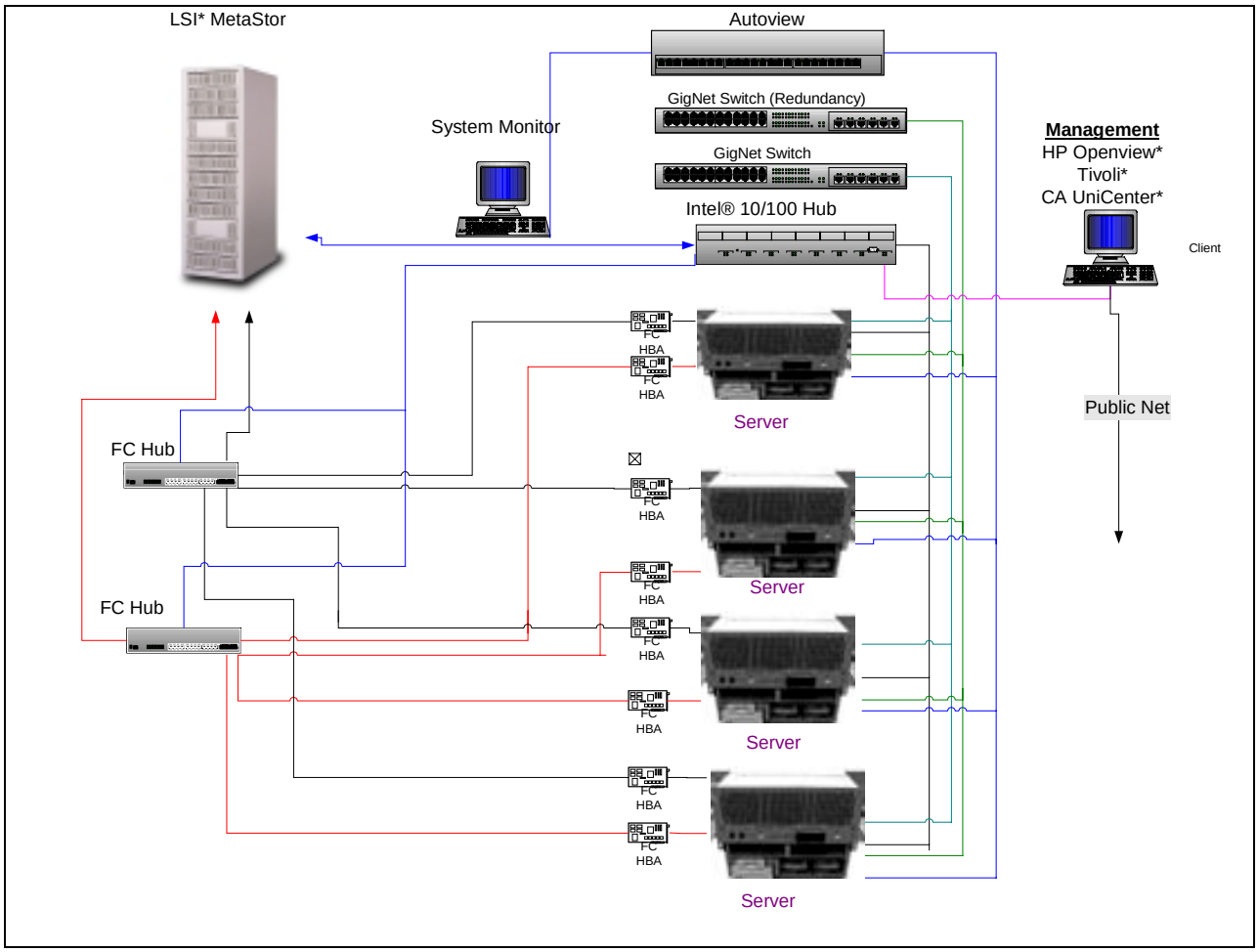

Figure 11*: *LSI Logic* MetaStor Oracle* Parallel Server Cluster

System Summary

The following table shows the equipment necessary to assemble the configuration.

Table 8: LSI Logic* Oracle* Parallel Server System Summary

Item	Model	Part #	Qty	Notes
Server: Intel Intel® AC450NX System, four 550-MHz Pentium® III Xeon™ processors, 1GB RAM **OR** Intel® OCPRF100 System, eight 550-MHz Pentium® III Xeon™ processors, 1GB RAM	AC450NX OCPRF100		4	System uses onboard SCSI drives to boot to Windows NT* 4.0 Enterprise Edition. Hot-plug PCI slots for “like for like“ replacement.
Monitor Switch: Cybex* Cybex Cables	AutoView Commander 8 ft monitor, mouse, keyboard	AV-8B CUFC-8	1 4	See Cybex for more info.
Storage: LSI Logic*				
72 inch Cabinet	Rack Mount	6256-4000-6600	1	
Rack Mount Rail Sets	DM/CM	6256-F958-0000	4	
Command Module	Rack for 4766	1000-3101-6600	1	
Controller	4766 256MB	1000-F120-0000	2	
Drive Module	Rack Mount	2000-2101-6600	3	
SCSI Connect CRU	LVD-SE	2000-F038-000	3	
Drives, 9GB U2	10K	2000-F210-0000	30	
Controller Code3.X	Software	1000-F951-0000	1	
Fiber Optic cable	MM	006-1086417	2	
FC HBA: Emulex* FW 2.81 Qlogic BIOS 1.35	 LP8000 Qla2100/66	 LP8000-N1 Qla2100/66-BK	8	
FC Hub: Emulex GBICs for Hub	LH5000 Optical	LH5000 LGB 100-N1	2 10	Managed through 10Mb Ethernet
Fiber Optic Cable: QFC	MM 3M	Q2xM2203	8	
VI HBA: Giganet	cLAN1000	cLAN1000	8	2 HBA per system
VI Switch: Giganet	cLAN5000	cLAN5000	2	
VI cables	2M HSSDC-HSSDC	cLAN-A0211	8	
Ethernet NIC: Intel	Intel® EtherExpress™ PRO/100+	689661-004	4	Ethernet 10/100Mbps
Ethernet HUB: Intel	Intel® 10/100 Stackable HUB	662197-010	1	Running at 100Mbs.
Ethernet CAT5 Cable	5 to 15 feet		7	Longer lengths may be needed
Operating system	Windows NT 4.0 SP5			Microsoft*
Data Base: Oracle*	Oracle Parallel Server 8.1			Contact Oracle for specific application needs
Enterprise management: Choose one-	HP OpenView* Tivoli* UniCenter*		1	Hewlett-Packard* Tivoli CAI*

3. Future Work and Trends

3.1 Storage Area Network

The basic idea behind SAN architecture is to share a common storage pool among many servers. SAN architecture dynamically allocates the storage resources among different servers without the need for physical configuration changes, thus avoiding down time. The main stream adoption of Fibre Channel has proven to be the most important enabling factor for SAN architecture.

The following diagrams illustrate the concept behind the SAN topology. The new technology that is used in this illustration is the HAL VI solution from Fujitsu*. The HAL VI is an interconnect that allows servers to communicate with one another at high speeds.

Additional technologies illustrated in this topology are FC hubs and switches that connect the storage to the servers along with a Fast Ethernet 1000G NIC, hubs and switches connecting with the requestor clients. A tape backup unit that is controlled by the Network Administrator is included as a back-up solution.

The advantage to this SAN technology solution is the ability to dynamically add/remove servers or storage components.

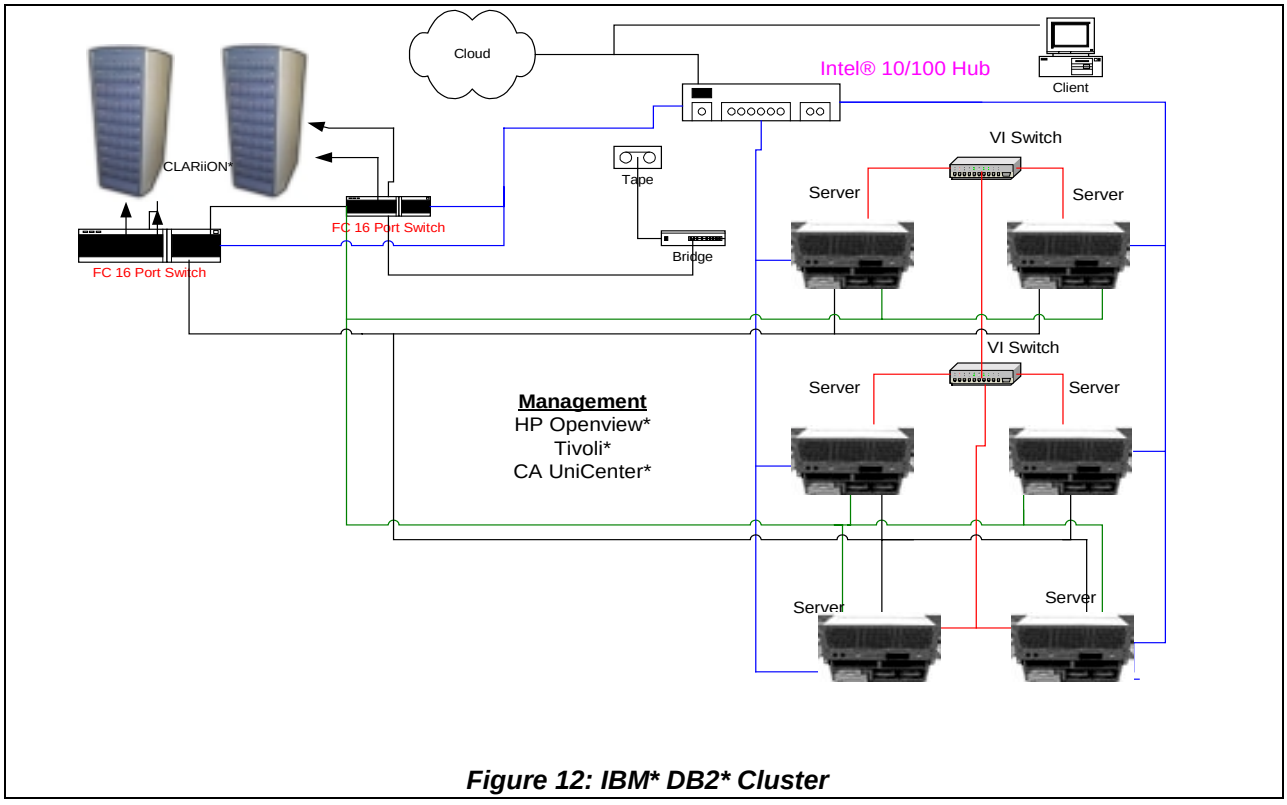

Figure 12: IBM* DB2* Cluster

IBM* DB2* is a cluster solution where performance is a factor. Six servers with Pentium III Xeon processors power this high performance DB SAN configuration. High Speed Interconnects such as the HAL VI will speed up requested transactions. The FC hubs and switches create the large bandwidths needed to forward the requests made of the server to the storage.

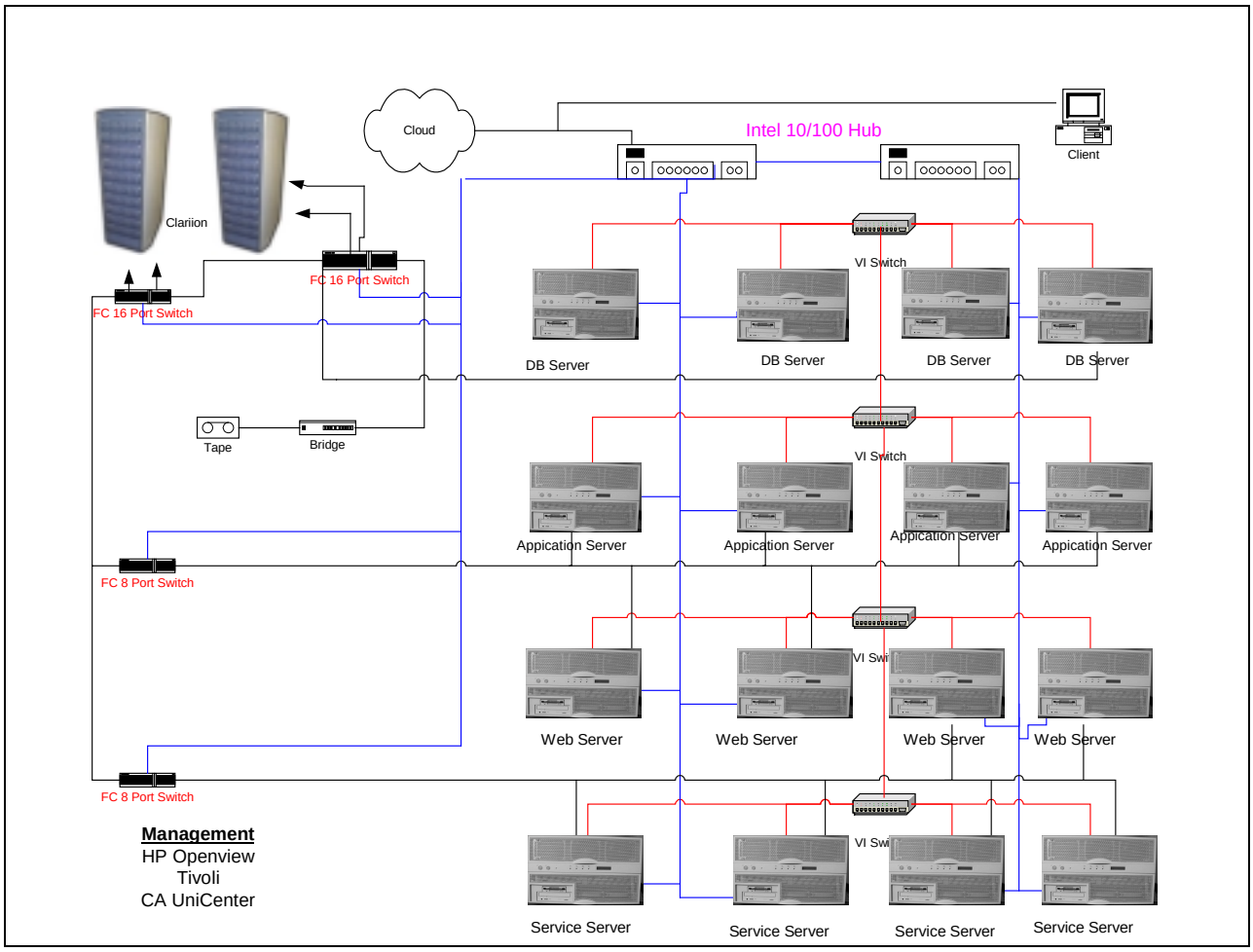

Figure 13: Three-tier Internet Cluster

The Internet cluster is another solution where performance is a factor. As with the DB cluster, Application, Web and Service Servers are all working together in a high performance SAN configuration. High Speed Interconnects such as the HAL VI will speed up requested transactions. The FC hubs and switches create the large bandwidths needed to forward the requests made of the server to the storage.

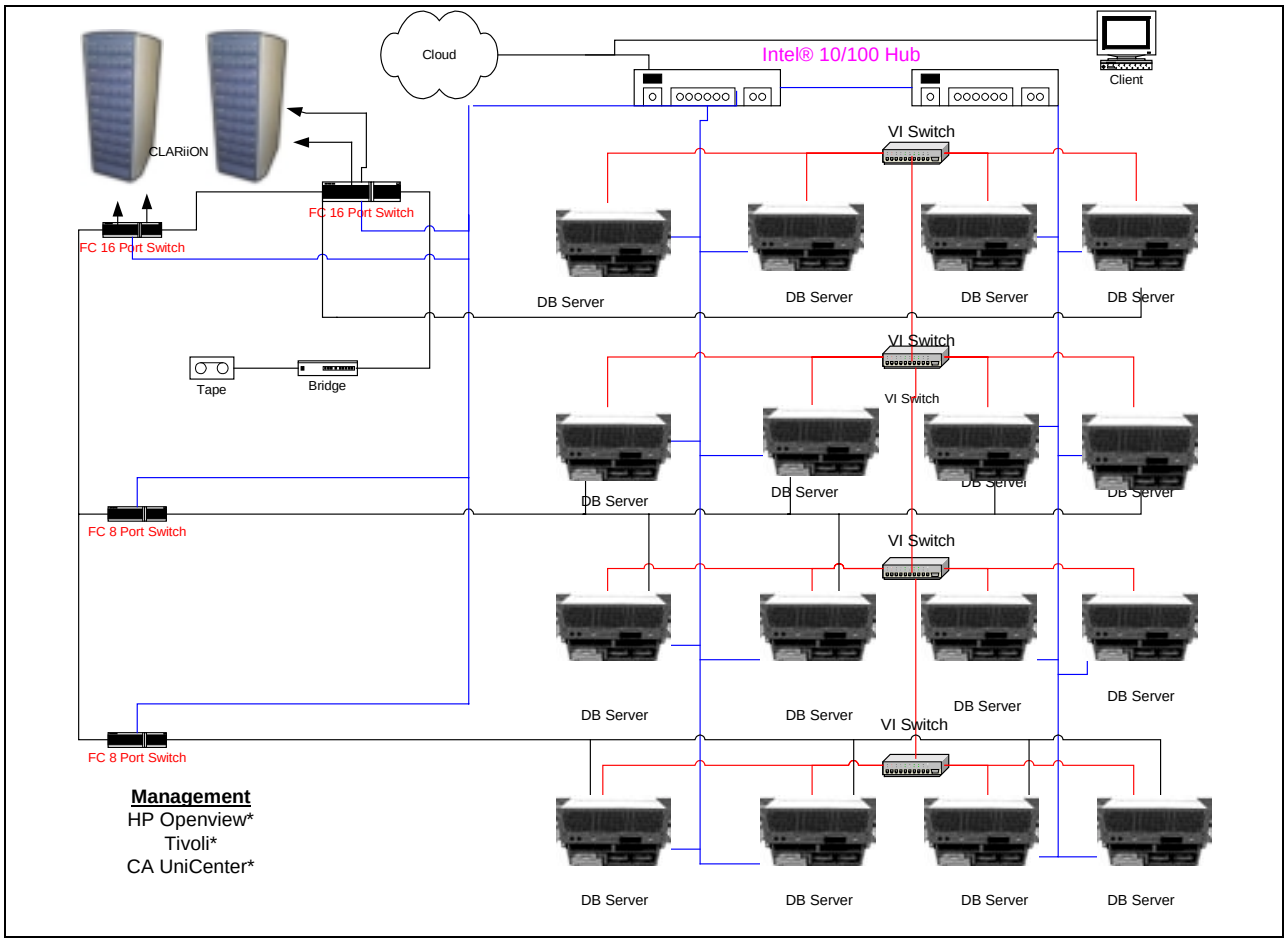

Figure 14: 16-Node DB2* Cluster

The 16-Node DB2 cluster is another solution where performance is a factor. The servers are all working together in a high performance SAN configuration. High Speed Interconnects such as the HAL VI will speed up requested transactions. The FC hubs and switches create the large bandwidths needed to forward the requests made of the clients to the storage.

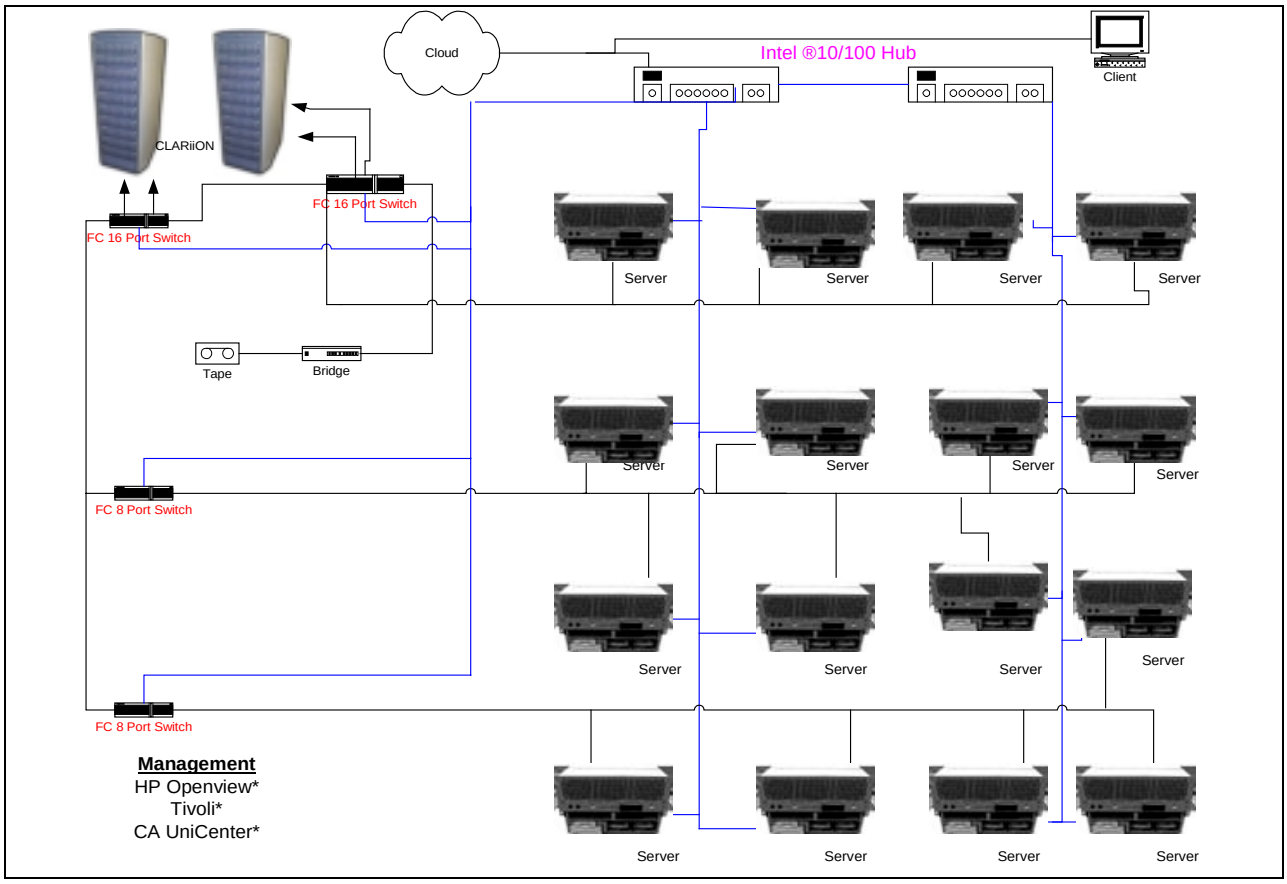

Figure 15: 16-Node Administrative Cluster

Administrative cluster is another solution where performance and not processor speed is a factor. The Administrative cluster will be a multi-application SAN configuration where the processors all work together in a high performance SAN configuration. The FC hubs and switches create the large bandwidths needed to forward the requests made of the clients to the storage.

4. Conclusion

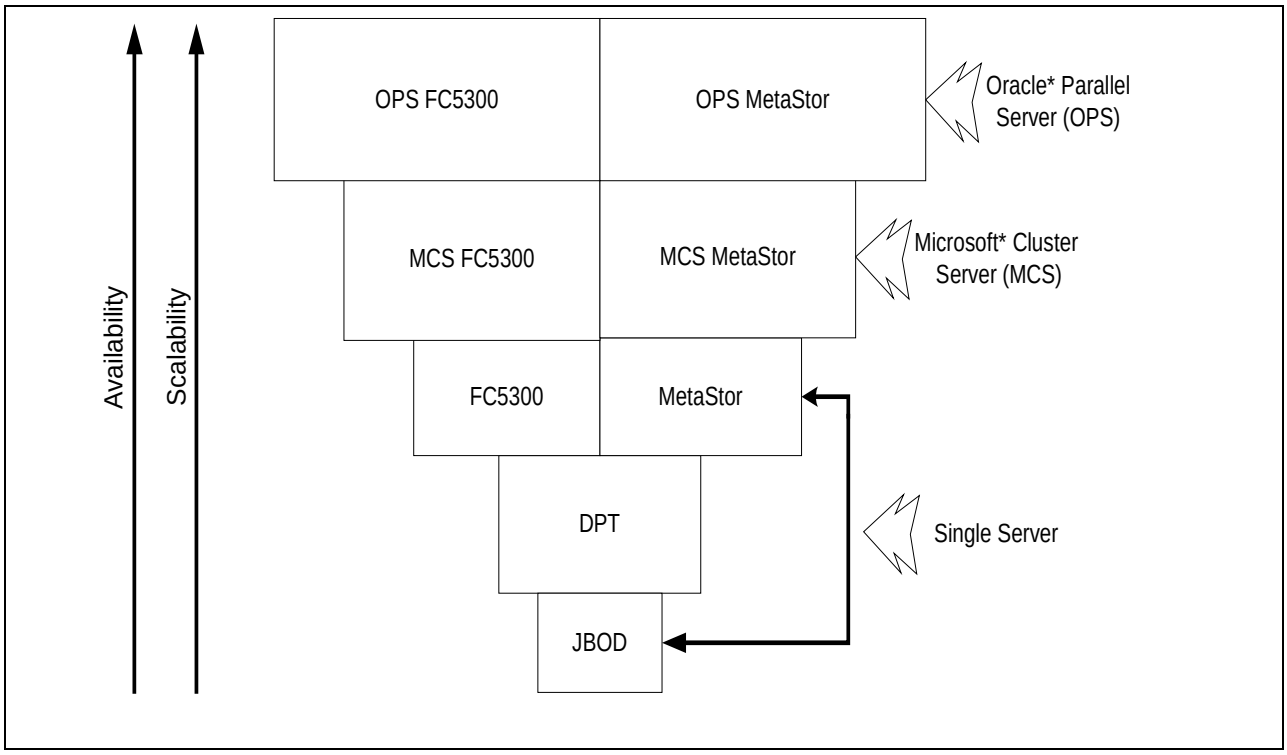

Figure 16: Storage Solutions

Enterprise Storage is a dynamically evolving issue. The ultimate storage goal is always being available and at the same time maintaining the flexibility to continually increase in scale. The server models shown here reflect real world, affordable scenarios.

JBOD accomplishes the goal of storing data. To insure we have a back-up of this data, RAIDs are utilized, increasing the availability of the data. When more availability is needed, larger RAIDs and redundant components are added. Eventually, redundant servers are added. Finally, applications with built in teaming, fail-over, and management take advantage of the latest storage technology, complete this scale.

This short overview demonstrates the maturity of clustering not only into the different price points but also into the many different segments of the computing world. Increased speed, bandwidth and performance are now available with the integration of high speed interconnects working in concert with FC hubs and switches. Increased scalability along with reduced cost is increasing the viability of SAN and cluster usage. Teaming with the power of systems with the Intel Pentium III Xeon processors enhance the flexibility and strength of today's SAN. These new solutions are expected to grow well into the next millennium.

Appendix A

This white paper will be updated as results become available. The results and updates will be listed in the appendices of this paper. The online address for this paper is http://support.intel.com/support/motherboards/server/ocprf100/s_storage.htm.

Currently the following sections of this paper are in validation:

- Section 2.1.4, LSI Logic System.
- Section 2.2.1, CLARiiON.
- Section 2.2.2, LSI Logic.
- Section 2.3.2, LSI Logic.