ExpressCluster for Linux Ver3.0

Operational Environment

2004.10.22
2st Revision

Revision History

Revision	Revision Date	Description
1	2004/06/30	New manual
2	2004/10/22	2.2.1.5 “About a File System on a Shared Disk” was added. 2.2.2.5 “About a File System on a Mirrored Disk” was added. 2.2.3.4 “About a File System on a Shared Disk” was added. 2.4 “OS and Network Interface that Operation has been Confirmed by Network Bonding setting” was added.

The latest information on system confirmation, system configuration guide, update, and tracking tool is provided in the following URL.
Please obtain the latest version before configuring the system.

Usage on the NEC Internet:
http://soreike.wsd.mt.nec.co.jp/

Usage out of the NEC Internet:
http://www.ace.comp.nec.co.jp/CLUSTERPRO/

1 ExpressCluster Versions Discussed in this Manual

This manual is written for the following ExpressCluster versions.

Keep in mind that some items, such as support function and kernel versions, are version-dependent.

Edition	Version
SE	3.0-4
LE	3.0-4
XE	3.0-1

2 Server Operational Environment

2.1 Hardware

The minimum spec requirements of servers are:

+ RS-232C Port 1 (not necessary when configuring a cluster of 3 or more nodes)
+ Ethernet Port 2 or more
+ FD Drive
+ CD-ROM Drive

Configure and connect the peripheral devices and network as shown in the following pages.

2.1.1 Operation Verified Disk Interface

The following disk types have been verified to run as LE (mirror data model) mirror disks.

Disk Type	Host side Driver name	Remarks
IDE	ide	Operation verified up to 120GB
SCSI	aic7xxx	
SCSI	sym53c8xx	
RAID	megaraid	

2.1.2 Operation Verified Network Interface

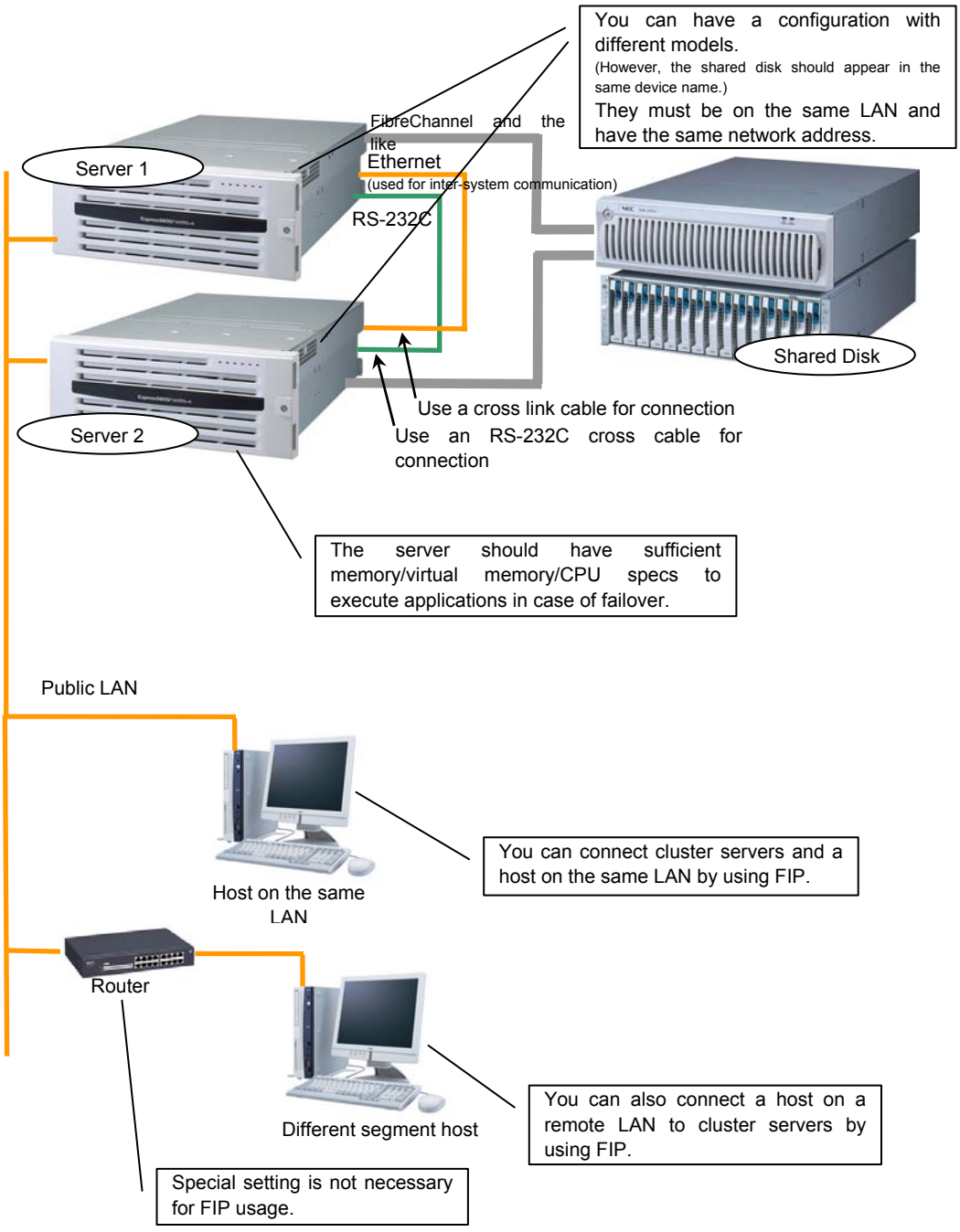
The following disk boards have been verified to run as mirror disk connect (systems for mirror communication) of data mirroring disks for LE (mirror data model).

Chip Name	Driver Name	Distribution[1]
Intel 82557/8/9	eepro100	TurboLinux ES8
	e100	RedHat ES3/AS3
Intel 82540EM	e1000	TurboLinux ES8 RedHat ES3/AS3
Intel 82544EI	e1000	RedHat ES3/AS3

[1] This does not mean a network interface does not operate on a board-driver combination with a distribution that is not listed here.

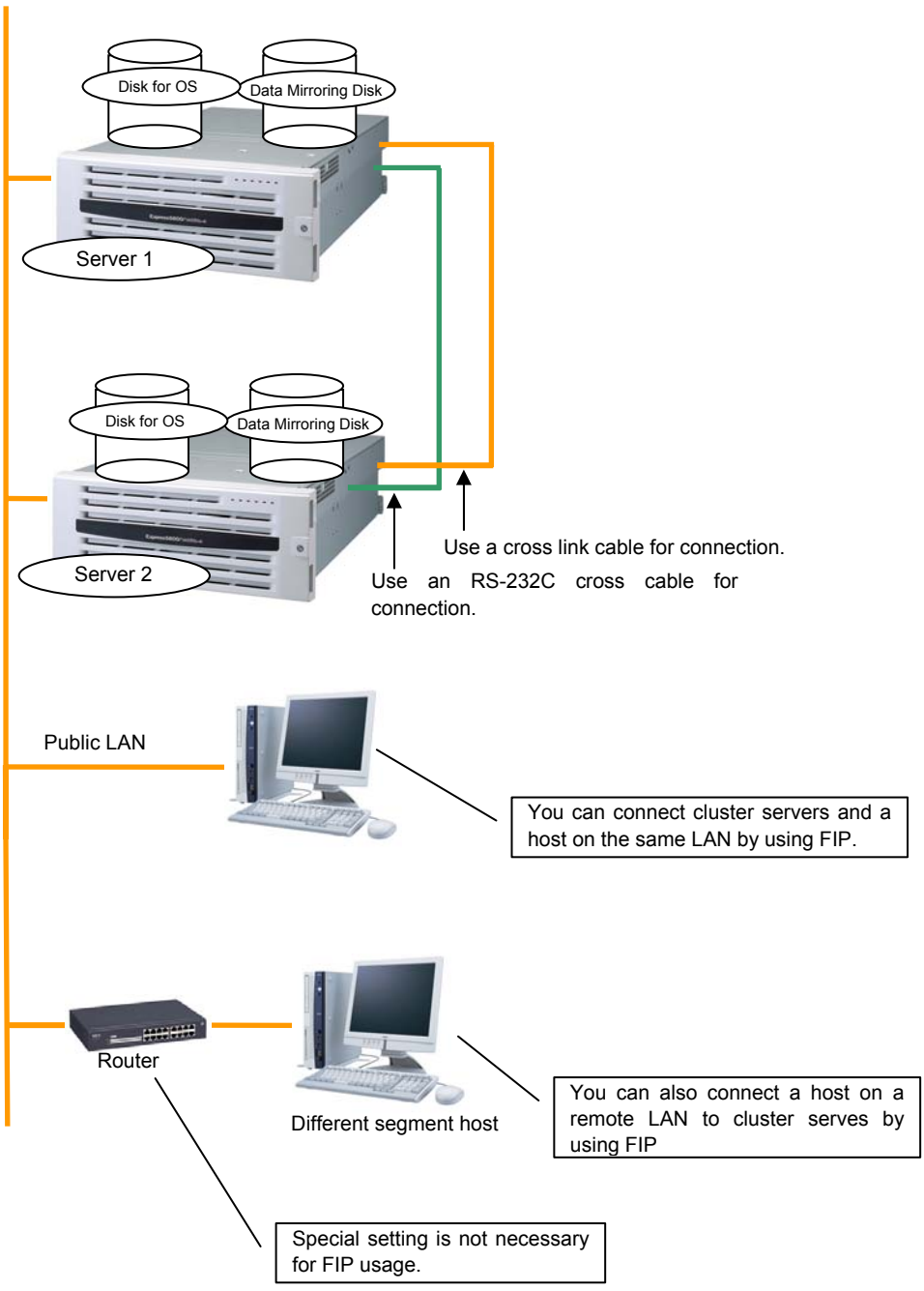
2.1.3 Example of SE Configuration with 2 Nodes

* You can use different models for servers. However, the shared disk should appear in the same device name in all servers.
* Connect the shared disk.
* Use cross cables for the interconnect connectivity. (You can use a dedicated HUB for connection just as the case of 4 Nodes.)
* Connect COM (RS-232C) ports using cross cables.



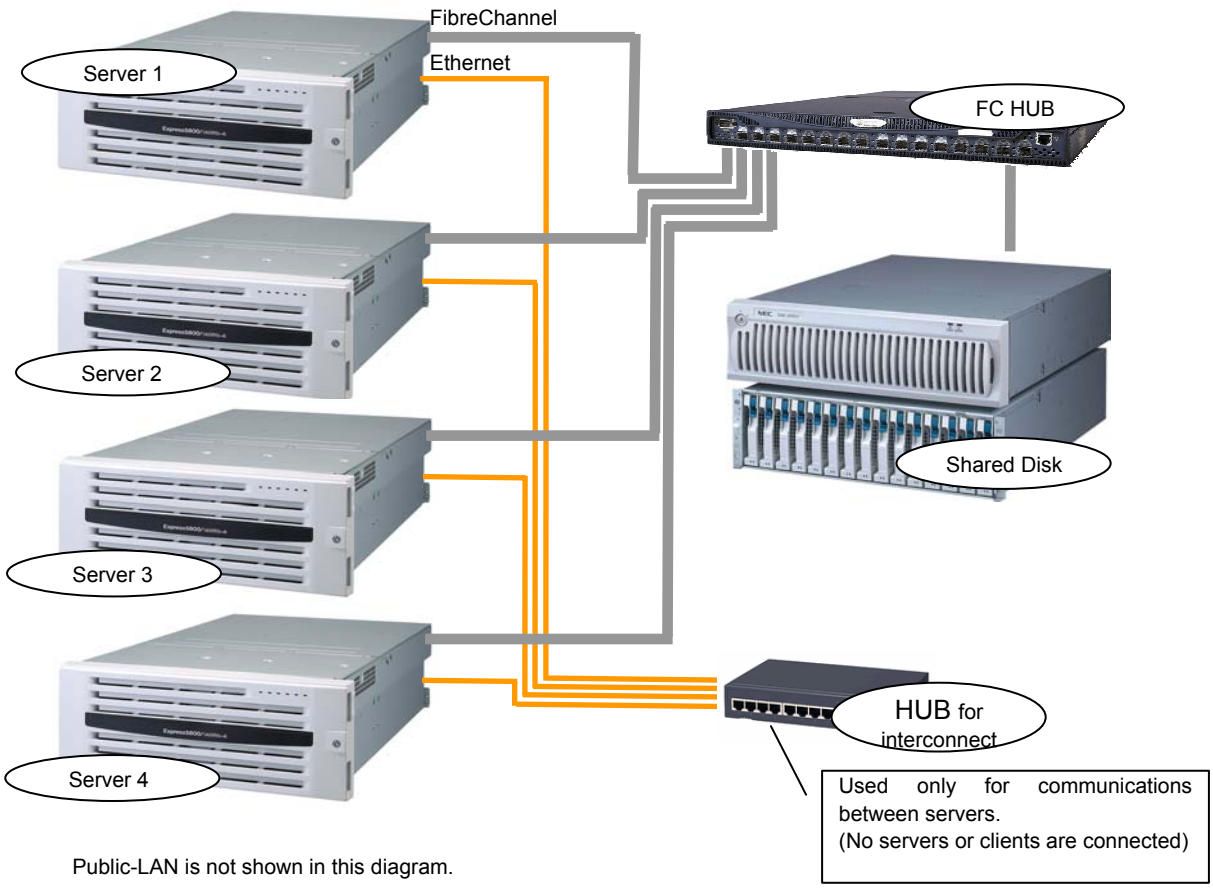
2.1.4 Example of LE Configuration with 2 Nodes

* You can use different models for servers. However, a mirroring disk should appear in the same device name in all servers.
* Use cross cables for the interconnect connectivity. Directly connect the interconnects for mirror (mirror disk connect)using cross cables. Do not connect a HUB or the like.
* Connect COM (RS-232C) ports using cross cables.



2.1.5 Example of SE Configuration with 4 Nodes

* As with the case of 2 nodes, the shared disk is connected. (The shared disk should appear in the same name in all servers.)
* Install a dedicated HUB for interconnect connectivity.
* You don't need to establish connectivity between servers using an RS-232C.

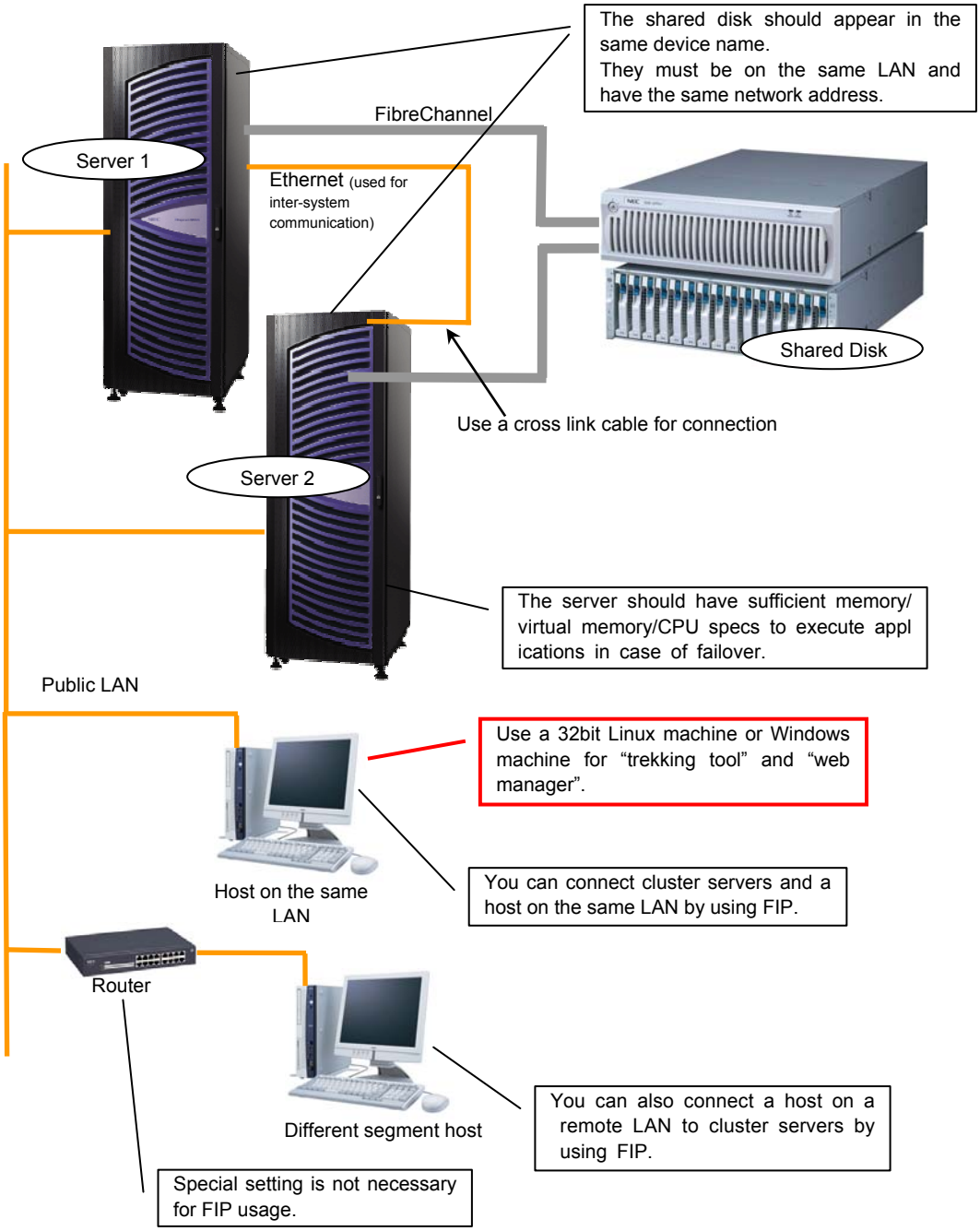


Public-LAN is not shown in this diagram.

2.1.6 Example of XE Configuration with 2 Nodes

The following is a conceptual diagram for TX-7. For information on connecting a shared disk, see related documents.

* The shared disk should appear in the same device name on all servers.
* Connect the shared disk.
* Use cross cables for interconnect connectivity. (You can use a dedicated HUB for connection just as the case of 4 Nodes.)



2.2 Software

2.2.1 SE --Shared Disk Model--

2.2.1.1 Operation Verified Distributions and Kernels

Since SE does not have its own kernel module, in principle, it does not depend on any kernel version. It has been verified SE runs with the following versions of kernels. With versions other than those listed below, SE may not operate properly.

The latest information is available on the ExpressCluster website.

Distribution	Kernel Version[2]	Binary RPM name of kernel
Turbolinux 8 Server	2.4.18-10 2.4.18-10smp 2.4.18-10smp64G	kernel-2.4.18-10.i586.rpm kernel-smp-2.4.18-10.i586.rpm kernel-smp64G-2.4.18-10.i586.rpm
	2.4.18-13 2.4.18-13smp 2.4.18-13smp64G	kernel-2.4.18-13.i586.rpm kernel-smp-2.4.18-13.i586.rpm kernel-smp64G-2.4.18-13.i586.rpm
Turbolinux Enterprise Server 8 powered by UnitedLinux	2.4.19-4GB 2.4.19-64GB-SMP	k_deflt-2.4.19-304.i586.rpm k_smp-2.4.19-304.i586.rpm
	2.4.19-4GB 2.4.19-64GB-SMP	k_deflt-2.4.19-333.i586.rpm k_smp-2.4.19-333.i586.rpm
	2.4.19-4GB 2.4.19-64GB-SMP	k_deflt-2.4.19-340.i586.rpm k_smp-2.4.19-340.i586.rpm
	2.4.21-138-default 2.4.21-138-smp	k_deflt-2.4.21-138.i586.rpm k_smp-2.4.21-138.i586.rpm
	2.4.21-198-default 2.4.21-198-smp	k_deflt-2.4.21-198.i586.rpm k_smp-2.4.21-198.i586.rpm
	2.4.21-215-default 2.4.21-215-smp	k_deflt-2.4.21-215.i586.rpm k_smp-2.4.21-215.i586.rpm
Red Hat Linux 7.2	2.4.20-18.7 2.4.20-18.7smp 2.4.20-18.7bigmen	kernel-2.4.20-18.7.i686.rpm kernel-smp-2.4.20-18.7.i686.rpm kernel-bigmen-2.4.20-18.7.i686.rpm
	2.4.20-20.7 2.4.20-20.7smp 2.4.20-20.7bigmen	kernel-2.4.20-20.7.i686.rpm kernel-smp-2.4.20-20.7.i686.rpm kernel-bigmen-2.4.20-20.7.i686.rpm
Red Hat Linux 7.3	2.4.20-18.7 2.4.20-18.7smp 2.4.20-18.7bigmen	kernel-2.4.20-18.7.i686.rpm kernel-smp-2.4.20-18.7.i686.rpm kernel-bigmen-2.4.20-18.7.i686.rpm
	2.4.20-20.7 2.4.20-20.7smp 2.4.20-20.7bigmen	kernel-2.4.20-20.7.i686.rpm kernel-smp-2.4.20-20.7.i686.rpm kernel-bigmen-2.4.20-20.7.i686.rpm
Red Hat Linux 8.0	2.4.20-18.8 2.4.20-18.8smp 2.4.20-18.8bigmem	kernel-2.4.20-18.8.i686.rpm kernel-smp-2.4.20-18.8.i686.rpm kernel-bigmem-2.4.20-18.8.i686.rpm

[2]Operations for standard configurations (kernel binary in the binary RPM name column) with each kernel installed have been verified. With kernels you built, SE may not operate.

Distribution	Kernel Version[2]	Binary RPM name of kernel
	2.4.20-20.8 2.4.20-20.8smp 2.4.20-20.8bigmen	kernel-2.4.20-20.8.i686.rpm kernel-smp-2.4.20-20.8.i686.rpm kernel-bigmen-2.4.20-20.8.i686.rpm
Red Hat Linux 9.0	2.4.20-6 2.4.20-6smp 2.4.20-6bigmen	kernel-2.4.20-6.i686.rpm kernel-smp-2.4.20-6.i686.rpm kernel-bigmem-2.4.20-6.i686.rpm
	2.4.20-20.9 2.4.20-20.9smp 2.4.20-20.9bigmen	kernel-2.4.20-20.9.i686.rpm kernel-smp-2.4.20-20.9.i686.rpm kernel-bigmen-2.4.20-20.9.i686.rpm
Red Hat Advanced Server2.1	2.4.9-e.10 2.4.9-e.10smp 2.4.9-e.10enterprise	kernel-2.4.9-e.10.i686.rpm kernel-smp-2.4.9-e.10.i686.rpm kernel-enterprise-2.4.9-e.10.i686.rpm
Red Hat Enterprise Linux AS/ES 2.1	2.4.9-e.16 2.4.9-e.16smp 2.4.9-e.16enterprise	kernel-2.4.9-e.16.i686.rpm kernel-smp-2.4.9-e.16.i686.rpm kernel-enterprise-2.4.9-e.16.i686.rpm
Red Hat Enterprise Linux AS/ES 3	2.4.21-4EL 2.4.21-4ELsmp 2.4.21-4ELhugemem	kernel-2.4.21-4.EL.i686.rpm kernel-smp-2.4.21-4.EL.i686.rpm kernel-hugemem-2.4.21-4.EL.i686.rpm
MIRACLE LINUX2.1	2.4.9-e.9.36ml 2.4.9-e.9.36mlsmp	kernel-2.4.9-e.9.36ml.i686.rpm kernel-smp-2.4.9-e.9.36ml.i686.rpm
MIRACLE LINUX V3.0	2.4.21-9.30AX 2.4.21-9.30AXsmp 2.4.21-9.30AXhugemem	kernel-2.4.21-9.30AX.i686.rpm kernel-smp-2.4.21-9.30AX.i686.rpm kernel-hugemem-2.4.21-9.30AX.i686.rpm

2.2.1.2 Depended Library

- libxml2

* Please install libxml2 when you install an OS.
For MIRACLE LINUX2.1, see the ExpressCluster website when installing libxml2.

2.2.1.3 Depended Driver

- softdog

* When you change a configuration for loadable modules, do not delete the driver shown above.

2.2.1.4 Required Memory Capacity and Disk Size

(1) Required Memory Capacity
- 22MB

(2) Disk Size
- Immediately after Installation: 8MB
- Maximum space during operation: 320MB*

* When you collect logs from all servers in the entire cluster using the clplogcc command, the maximum of 320 MB x the number of servers of free space will be necessary in a server that runs the command.

2.2.1.5 About a File System on a Shared Disk

The file system used on a shared disk is not dependant, but depending on the specification of the file system's fsck, problem may occur.

* The operation capability of the following file systems are currently confirmed.
- ext2
- ext3
- xfs
- reiserfs
- jfs
- vxfs

To decrease the chances of failure, it is recommended to use a file system with journal function.

2.2.2 LE --Data Mirror Model--

2.2.2.1 Operational Distribution and Kernel

Since LE has its own kernel module, it is dependent on kernel versions. The following are versions of kernels having modules (drivers) with which LE operates. LE does not operate properly with versions other than those listed below.

The latest information is available on the ExpressCluster website.

Distribution	Kernel Version[3]	Binary RPM name for kernel
Turbolinux 8 Server	2.4.18-16 2.4.18-16smp 2.4.18-16smp64G	kernel-2.4.18-16.i586.rpm kernel-smp-2.4.18-16.i586.rpm kernel-smp64G-2.4.18-16.i586.rpm
	2.4.18-17 2.4.18-17smp 2.4.18-17smp64G	kernel-2.4.18-17.i586.rpm kernel-smp-2.4.18-17.i586.rpm kernel-smp64G-2.4.18-17.i586.rpm
Turbolinux Enterprise Server 8 powered by UnitedLinux	2.4.19-4GB 2.4.19-64GB-SMP	k_deflt-2.4.19-340.i586.rpm k_smp-2.4.19-340.i586.rpm
	2.4.21-138-default 2.4.21-138-smp	k_deflt-2.4.21-138.i586.rpm k_smp-2.4.21-138.i586.rpm
	2.4.21-198-default 2.4.21-198-smp	k_deflt-2.4.21-198.i586.rpm k_smp-2.4.21-198.i586.rpm
	2.4.21-215-default 2.4.21-215-smp	k_deflt-2.4.21-215.i586.rpm k_smp-2.4.21-215.i586.rpm
Red Hat Enterprise Linux AS/ES 2.1	2.4.9-e.27 2.4.9-e.27smp 2.4.9-e.27enterprise	kernel-2.4.9-e.27.i686.rpm kernel-smp-2.4.9-e.27.i686.rpm kernel-enterprise-2.4.9-e.27.i686.rpm
	2.4.9-e.38 2.4.9-e.38smp 2.4.9-e.38enterprise	kernel-2.4.9-e.38.i686.rpm kernel-smp-2.4.9-e.38.i686.rpm kernel-enterprise-2.4.9-e.38.i686.rpm
Red Hat Enterprise Linux AS/ES 3	2.4.21-4EL 2.4.21-4ELsmp 2.4.21-4ELhugemem	kernel-2.4.21-4.EL.i686.rpm kernel-smp-2.4.21-4.EL.i686.rpm kernel-hugemem-2.4.21-4.EL.i686.rpm
	2.4.21-9.0.1EL 2.4.21-9.0.1ELsmp 2.4.21-9.0.1ELhugemem	kernel-2.4.21-9.0.1.EL.i686.rpm kernel-smp-2.4.21-9.0.1.EL.i686.rpm kernel-hugemem-2.4.21-9.0.1.EL.i686.rpm
MIRACLE LINUX2.1	2.4.9-e.25.36ml 2.4.9-e.25.36mlsmp	kernel-2.4.9-e.25.36ml.i686.rpm kernel-smp-2.4.9-e.25.36ml.i686.rpm
	2.4.9-e.25.50ml 2.4.9-e.25.50mlsmp	kernel-2.4.9-e.25.50ml.i686.rpm kernel-smp-2.4.9-e.25.50ml.i686.rpm
	2.4.9-e.25.52ml 2.4.9-e.25.52mlsmp	kernel-2.4.9-e.25.52ml.i686.rpm kernel-smp-2.4.9-e.25.52ml.i686.rpm
	2.4.9-e.25.60ml 2.4.9-e.25.60mlsmp 2.4.9-e.25.60mlenterprise	kernel-2.4.9-e.25.60ml.i686.rpm kernel-smp-2.4.9-e.25.60ml.i686.rpm kernel-enterprise-2.4.9-e.25.60ml.i686.rpm
	2.4.9-e.25.72ml 2.4.9-e.25.72mlsmp 2.4.9-e.25.72mlenterprise	kernel-2.4.9-e.25.72ml.i686.rpm kernel-smp-2.4.9-e.25.72ml.i686.rpm kernel-enterprise-2.4.9-e.25.72ml.i686.rpm

[3]Operations for standard configurations (kernel binary in the binary RPM name column) with each kernel installed have been verified. With kernels you built, LE may not operate.

Distribution	Kernel Version[3]	Binary RPM name for kernel
MIRACLE LINUX V3.0	2.4.21-9.30AX 2.4.21-9.30AXsmp 2.4.21-9.30AXhugemem	kernel-2.4.21-9.30AX.i686.rpm kernel-smp-2.4.21-9.30AX.i686.rpm kernel-hugemem-2.4.21-9.30AX.i686.rpm

2.2.2.2 Depended Library

- libxml2

* Please install libxml2 at the time of OS installation.
For MIRACLE LINUX2.1, see the ExpressCluster website when installing libxml2.

2.2.2.3 Depended Driver

- softdog

* When you change a configuration for loadable modules, do not delete the driver indicated above.

2.2.2.4 Required Memory Capacity and Disk Size

(1) Required Memory Capacity
Total of the following is necessary:

- 32MB (User Mode Process)
- 32MB + 2MB x the number of mirror sources (The area where kernel mode swap out is not possible)

Make sure to see a separate guide, “Trekking Tool” since parameters may require tuning, depending on the physical size of the memory, performance of CPU, and performance of the disk installed on the server, and the results of their operation.

(2) Disk Size
- Immediately after Installation: 8MB
- Maximum space during operation: 400MB*

* When you collect logs from all servers in the entire cluster using the clplogcc command, the maximum of 400MB x the number of servers of free space will be necessary in a server that runs the command.

2.2.2.5 About a File System on a Mirror Disk

The file system used on a mirror disk is not dependant, but depending on the specification of the file system's fsck, problem may occur.

* The operation capability of the following file systems are currently confirmed.
  - ext2
  - ext3
  - xfs

* The following file system is currently known to have a problem.
  - reiserfs (Because non-interactive fsck cannot be performed)

To decrease the chances of failure, it is recommended to use a file system with journal function.

2.2.3 XE --Shared Disk Model for TX-7--

2.2.3.1 Operation Verified Distribution and Kernel

Since XE does not have its own kernel module, it does not depend on any kernel versions in principle. It has been verified XE runs with the following version of kernel. With versions other than those listed below, XE may not operate properly.

The latest information is available on the ExpressCluster website.

Distribution	Kernel Version	Reference
Red Hat Advanced Server2.1	2.4.18-nec3.4p1.003	NEC kernel dedicated for TX-7

"Trekking Tool" and "Web Manager" do not operate on 64bit Linux.
You will need a 32bit Linux machine or Windows machine to construct and run the system.

2.2.3.2 Depended Library

- libxml2

2.2.3.3 Required Memory Capacity and Disk Size

(1) Required Memory Capacity
- 96MB

(2) Disk Size
- Immediately after Installation: 13MB
- Maximum space during operation: 320MB*

* When you collect logs from all servers in the entire cluster using the clplogcc command, the maximum free capacity of 320MB x the number of servers will be necessary in a server that runs the command.

2.2.3.4 About a File System on a Shared Disk

The file system used on a shared disk is not dependant, but depending on the specification of the file system's fsck, problem may occur.

* The operation capability of the following file systems are currently confirmed.
  - ext2
  - ext3
  - xfs

* The following file is currently known to have a problem.
  - reiserfs (Because non-interactive fsck cannot be performed)

To decrease the chances of failure, it is recommended to use a file system with journal function.

2.3 /opt/nec/clusterpro File Systems

* The operations of the following file systems have been verified:

    - ext2
    - ext3

To improve fault-tolerance of the system, it is recommended to use a file system with a journal function.

2.4 About Network Bonding Setting

See a separate guide, “Resource details” for more details.

The operation capability of the following distribution and network interface are currently confirmed.

2.4.1 Distribution

Distribution	Network Driver	Bonding Mode
RedHat ES/AS3 2.4.21-9.0.1.ELsmp	- bonding v2.2.14 - e100 2.3.30-k1 - e1000 5.2.20-k1	- active-backup(1) - balance-tlb(5)
TurboLinux ES8 2.4.21-231-smp	- bonding v2.2.14 - e100 2.3.27 - e1000 5.2.16	
MIRACLE LINUX V3.0 2.4.21-9.30AXsmp	- bonding v2.2.14 - e100 2.3.40 - e1000 5.2.39	

2.4.2 Network Interface

Ethernet Controller(Chip)	Bus	Driver
Intel 82557/8/9	PCI	e100
Intel 82544EI	PCI	e1000

Bonding setting of the network has been verified only by a public LAN.
Do not set it to interconnect or mirror connect.

3 Trekking Tool Operational Environment

3.1 Operation Verified OS

Linux (see 1.2.1 and 1.2.2. Trekking Tool does not operate on 64bit Linux)
Microsoft Windows® XP
Microsoft Windows® 2000

3.2 Operation Verified Browsers

* Browsers supporting Java 2

Mozilla 1.1 or later
Netscape® 7.1 or later
Microsoft® Internet Explorer 6.0 SP1 or later

3.3 Java Running Environments

* To use Trekking Tool, an environment where Java runs is necessary.

Sun Microsystems
Java(TM) 2 Runtime Environment, Standard Edition
Version 1.4.1_02 or later

3.4 Required Memory Capacity and Disk Size

Required Memory Capacity: Minimum of 32MB
Required Disk Space: 2MB

3.5 Combinations of OS and Browser Whose Operations Have Been Verified

The latest information is available on the ExpressCluster website.
At the time the product was released, the following environments had been verified to support Trekking Tool.

OS	Browser	Language
Windows® 2000 Professional	IE6 SP1	Japanese
Windows® 2000 Professional	Netscape 7.1	Japanese
Windows® XP Professional	IE6 SP1	Japanese
Windows® XP Professional	Netscape 7.1	English
Windows® XP Professional	IE6 SP1	English
Windows® 2000 Advanced Server	IE6 SP1	English
Turbolinux 8 Server	Mozilla 1.1	Japanese
Turbolinux 8 Enterprise Server	Netscape 7.1	English
Turbolinux 8 Enterprise Server	Netscape 7.1	Japanese
Red Hat Linux 9.0	Mozilla 1.2	English
Windows® 2000 Professional	Netscape 7.1	Japanese
Turbolinux 8 Enterprise Server	Mozilla 1.0.1	English
Turbolinux 8 Enterprise Server	Mozilla 1.0.1	Japanese
Red Hat Enterprise Linux ES 3.0	Mozilla 1.4	English
Red Hat Enterprise Linux AS 3.0	Mozilla 1.4	Japanese

3.6 Supported ExpressCluster Versions

--For SE--

Trekking Tool Version	ExpressCluster rpm Version
3.0-1	3.0-1 3.0-2
3.0-2 3.0-3	3.0-3

--For LE--

Trekking Tool Version	ExpressCluster rpm Version
3.0-2	3.0-1
3.0-3	3.0-2 3.0-3

--For XE--

Trekking Tool Version	ExpressCluster rpm Version
3.0-3	3.0-1

4 Web Manager Operational Environment

4.1 Operation Verified OS

Linux (see 1.2.1 and 1.2.2. Web Manager does not operate on 64bit Linux)
Microsoft Windows® XP
Microsoft Windows® 2000

4.2 Operation Verified Browsers

* Browsers supporting Java 2

Mozilla 1.1 or later
Netscape® 7.1 or later
Microsoft® Internet Explorer 6.0 SP1 or later

4.3 Java Running Environments

* To use Web manager, an environment where Java runs is necessary.

Sun Microsystems
Java(TM) 2 Runtime Environment, Standard Edition
Version 1.4.1_02 or later

4.4 Required Memory Capacity and Disk Space

Required Memory Capacity: Minimum of 40MB
Required Disk Capacity: 300KB (excluding Java running environment)

4.5 Combinations of OS and Browser Whose Operations Have Been Verified

The latest information is available on the ExpressCluster website.
At the time the product was released, the following environments had been verified to support Web Manager.

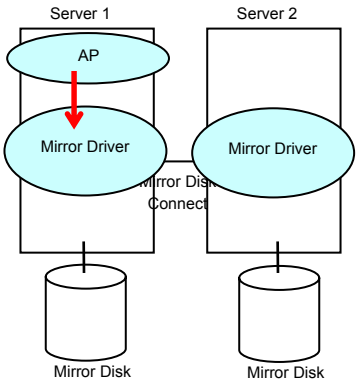
OS	Browser	Language
Windows® 2000 Professional	IE6 SP1	Japanese
Windows® 2000 Professional	Netscape 7.1	Japanese
Windows® XP Professional	IE6 SP1	Japanese
Windows® XP Professional	Netscape 7.1	English
Windows® XP Professional	IE6 SP1	English
Windows® 2000 Advanced Server	IE6 SP1	English
Turbolinux 8 Server	Mozilla 1.1	Japanese
Turbolinux 8 Enterprise Server	Netscape 7.1	English
Turbolinux 8 Enterprise Server	Netscape 7.1	Japanese
Red Hat Linux 9.0	Mozilla 1.2	English
Windows® 2000 Professional	Netscape 7.1	Japanese
Turbolinux 8 Enterprise Server	Mozilla 1.0.1	English
Turbolinux 8 Enterprise Server	Mozilla 1.0.1	Japanese
Red Hat Enterprise Linux ES 3.0	Mozilla 1.4	English

5 What You Should Know About LE

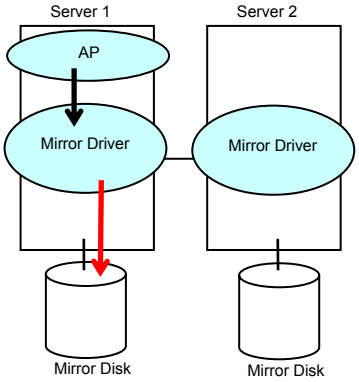
5.1 Mechanism of Mirroring

The following figure briefly shows the mechanism of LE mirroring through a typical write process.

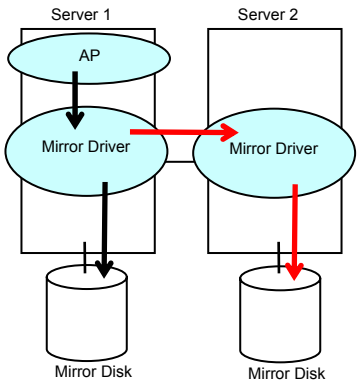
In actual operation, speedups by queing and retry also occur.



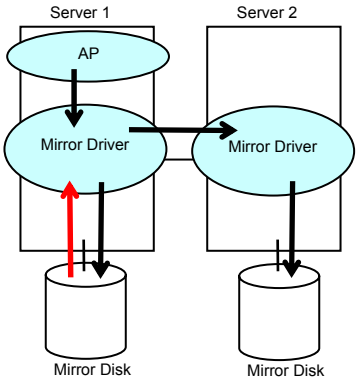
A write request from AP enters a mirror driver through a file system.
The file system waits for an I/O completion notice.
(File system layers are not shown in the figure.)



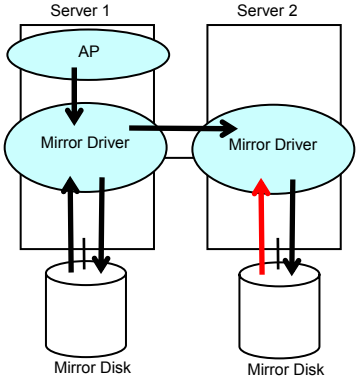
The write request which entered the mirror driver issues a write request to the active disk (block device layer).



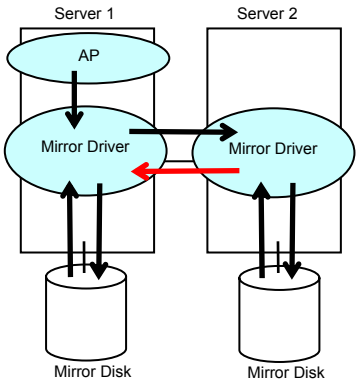
The active mirror driver transmit the write request to the stand-by mirror driver through a mirror connect without waiting for the IO completion.
The write request received by the stand-by mirror driver issues a write request to the stand-by disk (block device layer).



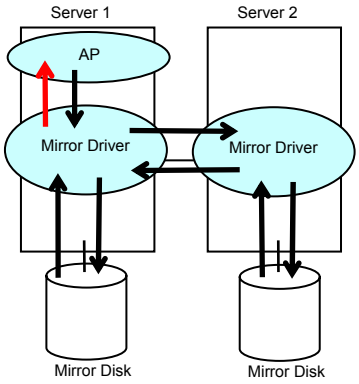
The active mirror driver recognizes the completion notice from OS.



The stand-by mirror driver recognizes completion notice from OS.
The sequence of the completion notice from the active/stand-by OS is not defined.



The stand-by mirror driver sends the write completion notice (ACK) to the active mirror driver.

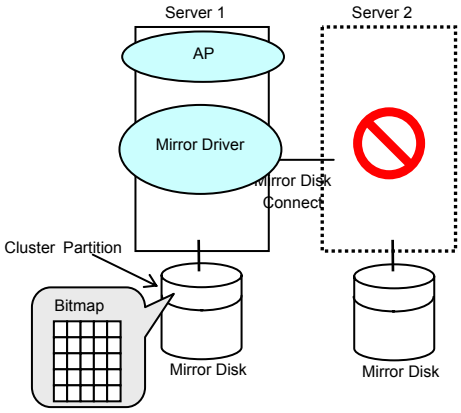


The active mirror driver notifies the file system of I/O completion.
(File system layers are not shown in the figure.)

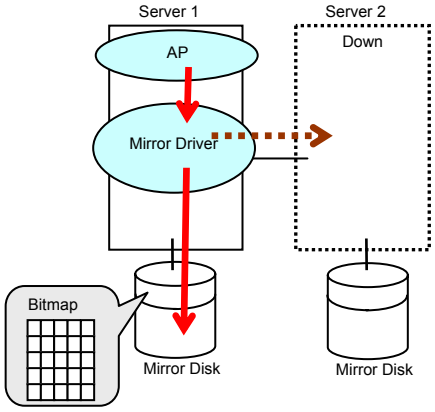
5.2 Mechanism of FastSync Option

The following figure briefly shows the mechanism how re-synchronization (mirror recovery) speed is improved when LE and FastSync options are used together.

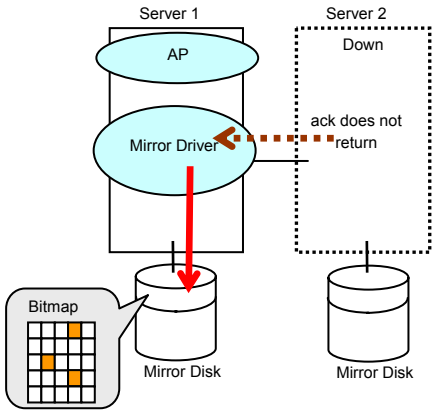
In actual operation, differential mirror recovery may not be achieved even FastSync Option is used. See a separate guide, “Maintenance” for conditions allowing differential mirror recovery.



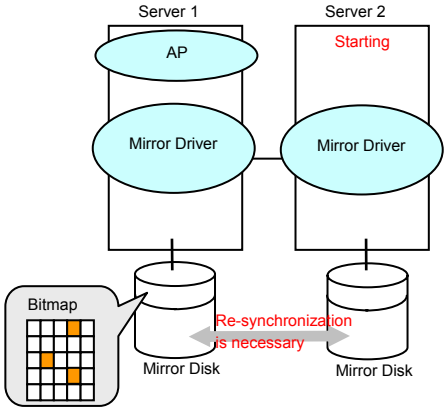
The bitmap which corresponds to the data partition block is stored in a cluster partition.



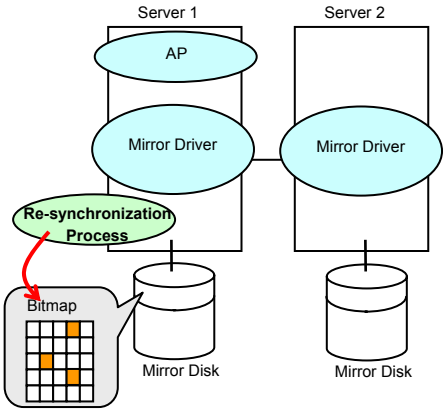
If the mirror disk in the working server is updated while the other server is failing, differences between the failed and working servers will be generated.



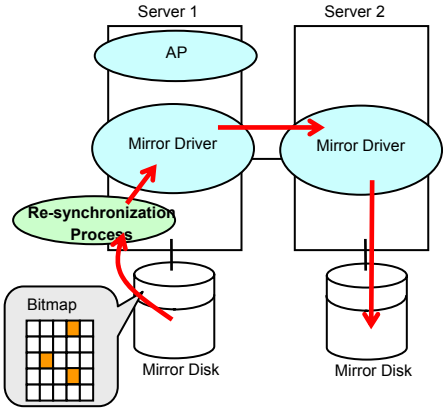
As for a write i/o without write response (ack), the bitmap is marked as “updated” with the block position as a key.



When the failed server starts up, re-synchronization (mirror recovery) of mirror disk data needs to be performed.



The information on the block requiring re-synchronization is obtained from the bitmap when the re-synchronization process begins.
The entire partition is re-synchronized (copied) if there is no FastSync Option.



The block data obtained from the bitmap is copied through mirror driver.

5.3 Precautions on Mirroring System

Make sure to read the following precautions and have good knowledge of the mirroring system before you make a proposal or establish a system.

5.3.1 About Network for Mirroring Disk Connect

* It is recommended to use high-speed network, such as 100Base-TX and 1000Base-T, for mirror disk connect.
* See 4.3.5 “Reference Performance Data” for the performance differences between 100Base-TX and 1000Base-T.

5.3.2 About Data Mirroring Disk

* You cannot use a disk which is controlled by OS (with a swap partition, or with a partition which is mounted and unmounted by /etc/fstab) as a data mirroring disk. Add one or more data mirroring disks on each server.

* Reserve a logical disk (how it is represented, such as “pack,” “logical unit,” “LUN,” varies depending on hardware vendor) for mirror for hardware RAID.
* If a Linux OS has been preinstalled and there is no free space (i.e. no space for new LUN) in a hardware RAID configuration, you may need to reconfigure the hardware RAID or reinstall the Linux OS.

* It is recommended to use a high-speed disk possible as a data mirroring disk.

* It is recommended to set the RAID board cache to WRITE BACK (or WRITE cache enabled) when using a hardware RAID.
  (Make sure to check RAID board specifications and precautions since and additional battery to back up the RAID BOARD memory or UPS may be necessary.)

5.3.3 About Write Performance of Mirror

* In the write process of data mirroring disks, data is written on remote server disks through network, and local server disks. In the reading process, only data on the local server disks is read.
* Thus, the write performance degrades when compared to single server.
  For a system requiring as high throughput as a single server for the writing process, such as a database system with frequent updates), the SE version should be proposed.
* See 4.3.5 “Reference Performance Data” as well for a write performance of mirror.

5.3.4 About Mirror Recover Time

* When one server goes down and then starts, data mirroring disk re-synchronization becomes necessary.
* Always copy the entire partitions if the FastSync option is not used.
* When FastSync Option is used, the recovery time may be shortened since only data differences are copied.
  However, depending on the condition of data mirroring disk, the entire partitions must be copied even when using FastSync Option. See a separate guide, “Maintenance” for details.
* Note that if mirror resources are heavily loaded during recovery (performing mirror recovery while transactions are continued), it will require an extra time and degrade performance while recovering.

* See 4.3.5 “Reference Performance Data” as well for mirror recovery time.

5.3.5 Reference Performance Data

5.3.5.1 100Base-TX and 1000Base-T Comparison

The following is a result of file copy (from a local disk to a data mirroring disk) using 100Base-TX and 1000Base-T on the same machine.

The following data does not guarantee the absolute data because the values shown here may vary according to CPU specs and disk performance. Use this data as a reference to a performance difference of 100Base-TX and 1000Base-T.

	From the start of cp command to the end of cp command	From the start of the cp command to the end of the sync command followed by cp command
100Base-TX (e100.o)	23MB/Sec	5.0MB/Sec
1000Base-T (e1000.o)	42MB/Sec	5.5MB/Sec
Local Disk (Reference Value)	-	6.0MB/Sec

[Measuring Condition]

Distribution	TurboLinux EnterpriseServer8
Kernel	kernel-2.4.19-340
Memory	1GB
CPU	Intel(R) Xeon(TM) 2.40GHz x 2
Disk Type	MegaRaid RAID5 (write cache off)
File System	ext3
Mount Option	rw
File Size	500MB

5.3.5.2 Write Performance Comparison

The following is a result of file copy and bonnie (a measurement tool for file read/write throughput) on the same machine.

The following data does not guarantee the absolute data because the values shown here may vary according to CPU specs, disk performance, network performance, data size of write, and file system. Use this data as a reference to a performance difference of a local disk and data mirroring disk.

[2GB file copy]

	From the start of cp command to the end of cp command	From the start of the cp command to the end of the sync command followed by cp command
From a local disk to a local disk	6.7MB/Sec	5.2MB/Sec
From a local disk to a data mirroring disk (Mirror disk connect 1000Base-T)	6.4MB/Sec	4.9MB/Sec

[Measurement Environment]

Distribution	TurboLinux EnterpriseServer8
Kernel	kernel-2.4.19-340
Memory	1GB
CPU	Intel(R) Xeon(TM) 2.40GHz x 2
Disk Type	MegaRaid RAID5 (write cache off)
File System	ext3
Mount Option	rw

[2047MB bonnie]

	Sequential Output (nosync) K/Sec		
	-Per Char-	-Block-	-Rewrite-
Local Disk	20105	34857	14253
Data mirroring disk (Mirror disk connect 100Base-TX)	13039	14294	10251

[Measurement Environment]

Distribution	TurboLinux EnterpriseServer8
Kernel	kernel-2.4.19-340
Memory	1.5GB
CPU	Intel(R) Pentium(R) III 1.40GHz x 2
Disk Type	SCSI (aic7xxx U160)
File System	ext3
Mount Option	rw

5.3.5.3 Mirror Recovery Time

The following is a result of mirror recovery time.

The following data does not guarantee the absolute data because the values shown here may vary according to CPU specs, disk performance, and network performance.

	Mirror Recovery Speed
Condition 1	6.6MB/Sec (Approx. 2.5 Min/GB)
Condition 2	2.8MB/Sec (Approx. 4.7 Min/GB)

[Condition 1]

Distribution	TurboLinux EnterpriseServer8
Kernel	kernel-2.4.19-340
Memory	1.5GB
CPU	Intel(R) Pentium(R) III 1.40GHz x 2
Disk Type	SCSI (aic7xxx U160)
Partition Size	4GB
Interconnect	100Base-TX

[Condition 2]

Distribution	TurboLinux EnterpriseServer8
Kernel	kernel-2.4.19-340
Memory	1GB
CPU	Intel(R) Xeon(TM) 2.40GHz x 2
Disk Type	MegaRaid RAID5 (write cache off)
Partition Size	4GB
Interconnect	1000Base-T