EXPRESSCLUSTER X 5.0 for Linux
Getting Started Guide

Release 4

NEC Corporation

Nov 04, 2022

TABLE OF CONTENTS:

CHAPTER

ONE

PREFACE

1.1 Who Should Use This Guide

EXPRESSCLUSTER X Getting Started Guide is intended for first-time users of the EXPRESSCLUSTER. The guide covers topics such as product overview of the EXPRESSCLUSTER, how the cluster system is installed, and the summary of other available guides. In addition, latest system requirements and restrictions are described.

1.2 How This Guide is Organized

- 2. *What is a cluster system?*: Helps you to understand the overview of the cluster system and EXPRESSCLUSTER.
- 3. *Using EXPRESSCLUSTER*: Provides instructions on how to use a cluster system and other related-information.
- 4. *Installation requirements for EXPRESSCLUSTER*: Provides the latest information that needs to be verified before starting to use EXPRESSCLUSTER.
- 5. *Latest version information*: Provides information on latest version of the EXPRESSCLUSTER.
- 6. *Notes and Restrictions*: Provides information on known problems and restrictions.
- 7. *Upgrading EXPRESSCLUSTER*: Provides instructions on how to update the EXPRESSCLUSTER.

1.3 EXPRESSCLUSTER X Documentation Set

The EXPRESSCLUSTER X manuals consist of the following five guides. The title and purpose of each guide is described below:

EXPRESSCLUSTER X Getting Started Guide

> This guide is intended for all users. The guide covers topics such as product overview, system requirements, and known problems.

EXPRESSCLUSTER X Installation and Configuration Guide

> This guide is intended for system engineers and administrators who want to build, operate, and maintain a cluster system. Instructions for designing, installing, and configuring a cluster system with EXPRESSCLUSTER are covered in this guide.

EXPRESSCLUSTER X Reference Guide

> This guide is intended for system administrators. The guide covers topics such as how to operate EXPRESSCLUSTER, function of each module and troubleshooting. The guide is supplement to the Installation and Configuration Guide.

EXPRESSCLUSTER X Maintenance Guide

> This guide is intended for administrators and for system administrators who want to build, operate, and maintain EXPRESSCLUSTER-based cluster systems. The guide describes maintenance-related topics for EXPRESSCLUSTER.

EXPRESSCLUSTER X Hardware Feature Guide

> This guide is intended for administrators and for system engineers who want to build EXPRESSCLUSTER-based cluster systems. The guide describes features to work with specific hardware, serving as a supplement to the Installation and Configuration Guide.

1.4 Conventions

In this guide, **Note**, **Important**, **See also** are used as follows:

Note: Used when the information given is important, but not related to the data loss and damage to the system and machine.

Important: Used when the information given is necessary to avoid the data loss and damage to the system and machine.

See also:

Used to describe the location of the information given at the reference destination.

The following conventions are used in this guide.

Convention	Usage	Example
Bold	Indicates graphical objects, such as fields, list boxes, menu selections, buttons, labels, icons, etc.	In User Name, type your name. On the File menu, click Open Database.
Angled bracket within the command line	Indicates that the value specified inside of the angled bracket can be omitted.	clpstat -s[-h host_name]
#	Prompt to indicate that a Linux user has logged on as root user.	# clpcl -s -a
Monospace	Indicates path names, commands, system output (message, prompt, etc.), directory, file names, functions and parameters.	`/Linux/5.0/en/server/`
bold	Indicates the value that a user actually enters from a command line.	Enter the following: **# clpcl -s -a**
italic	Indicates that users should replace italicized part with values that they are actually working with.	`clpstat -s [-h` *`host_name`*`]`

In the figures of this guide, this icon represents EXPRESSCLUSTER.

1.5 Contacting NEC

For the latest product information, visit our website below:

https://www.nec.com/global/prod/expresscluster/

CHAPTER

TWO

WHAT IS A CLUSTER SYSTEM?

This chapter describes overview of the cluster system.

This chapter covers:

2.1 Overview of the cluster system

A key to success in today's computerized world is to provide services without them stopping. A single machine down due to a failure or overload can stop entire services you provide with customers. This will not only result in enormous damage but also in loss of credibility you once enjoyed.

A cluster system is a solution to tackle such a disaster. Introducing a cluster system allows you to minimize the period during which operation of your system stops (down time) or to avoid system-down by load distribution.

As the word "cluster" represents, a cluster system is a system aiming to increase reliability and performance by clustering a group (or groups) of multiple computers. There are various types of cluster systems, which can be classified into the following three listed below. EXPRESSCLUSTER is categorized as a high availability cluster.

- **High Availability (HA) Cluster**

  In this cluster configuration, one server operates as an active server. When the active server fails, a standby server takes over the operation. This cluster configuration aims for high-availability and allows data to be inherited as well. The high availability cluster is available in the shared disk type, data mirror type or remote cluster type.

- **Load Distribution Cluster**

  This is a cluster configuration where requests from clients are allocated to load-distribution hosts according to appropriate load distribution rules. This cluster configuration aims for high scalability. Generally, data cannot be taken over. The load distribution cluster is available in a load balance type or parallel database type.

- **High Performance Computing (HPC) Cluster**

  This is a cluster configuration where CPUs of all nodes are used to perform a single operation. This cluster configuration aims for high performance but does not provide general versatility.

  Grid computing, which is one of the types of high performance computing that clusters a wider range of nodes and computing clusters, is a hot topic these days.

2.2 High Availability (HA) cluster

To enhance the availability of a system, it is generally considered that having redundancy for components of the system and eliminating a single point of failure is important. "Single point of failure" is a weakness of having a single computer component (hardware component) in the system. If the component fails, it will cause interruption of services. The high availability (HA) cluster is a cluster system that minimizes the time during which the system is stopped and increases operational availability by establishing redundancy with multiple servers.

The HA cluster is called for in mission-critical systems where downtime is fatal. The HA cluster can be divided into two types: shared disk type and data mirror type. The explanation for each type is provided below.

2.2.1 Shared disk type

Data must be inherited from one server to another in cluster systems. A cluster topology where data is stored in a shared disk with two or more servers using the data is called shared disk type.

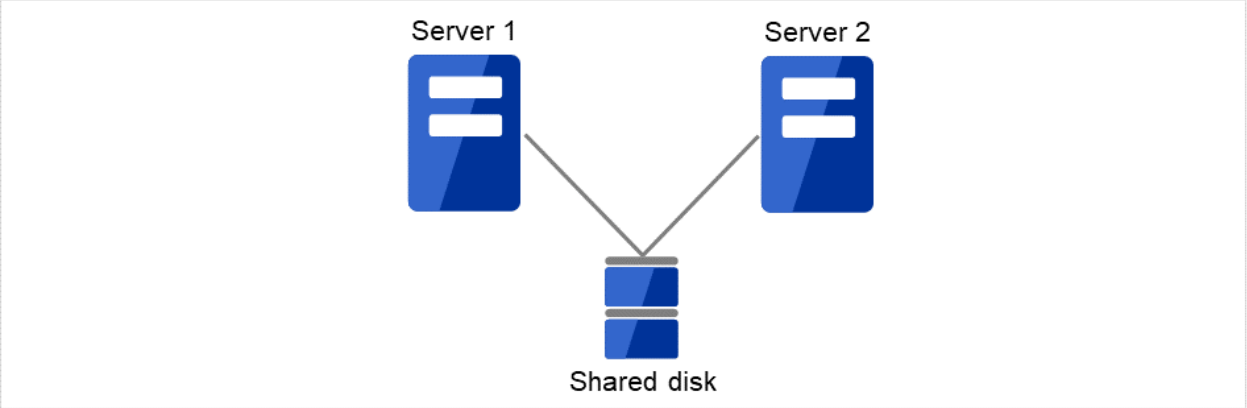


Fig. 2.1: HA cluster configuration (Shared disk type)

- Expensive since a shared disk is necessary.
- Ideal for the system that handles large data.

If a failure occurs on a server where applications are running (active server), the cluster system detects the failure and applications are automatically started in a standby server to take over operations. This mechanism is called failover. Operations to be inherited in the cluster system consist of resources including disk, IP address and application.

In a non-clustered system, a client needs to access a different IP address if an application is restarted on a server other than the server where the application was originally running. In contrast, many cluster systems allocate a virtual IP address on an operational basis. A server where the operation is running, be it an active or a standby server, remains transparent to a client. The operation is continued as if it has been running on the same server.

File system consistency must be checked to inherit data. A check command (for example, fsck in Linux) is generally run to check file system consistency. However, the larger the file system is, the more time spent for checking. While checking is in process, operations are stopped. For this problem, journaling file system is introduced to reduce the time required for failover.

Logic of the data to be inherited must be checked for applications. For example, roll-back or roll-forward is necessary for databases. With these actions, a client can continue operation only by re-executing the SQL statement that has not been committed yet.

A server with the failure can return to the cluster system as a standby server if it is physically separated from the system, fixed, and then succeeds to connect the system. Such returning is acceptable in production environments where continuity of operations is important.

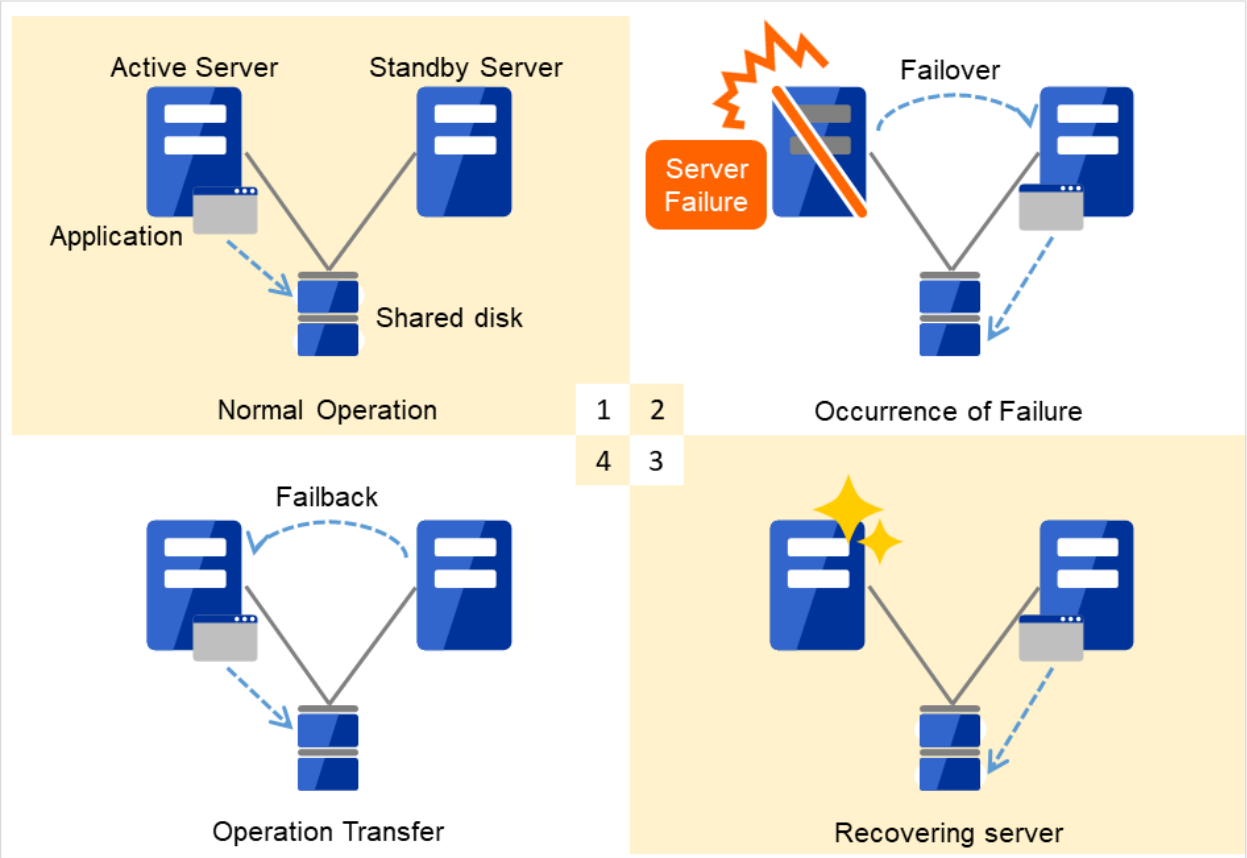


Fig. 2.2: From occurrence of a failure to recovery

1. Normal operation
2. Occurrence of failure
3. Recovering server
4. Operation transfer

When the specification of the failover destination server does not meet the system requirements or overload occurs due to multi-directional standby, operations on the original server are preferred. In such a case, a failback takes place to resume operations on the original server.

A standby mode where there is one operation and no operation is active on the standby server, as shown in Figure 2.3 From occurrence of a failure to recovery, is referred to as uni-directional standby.

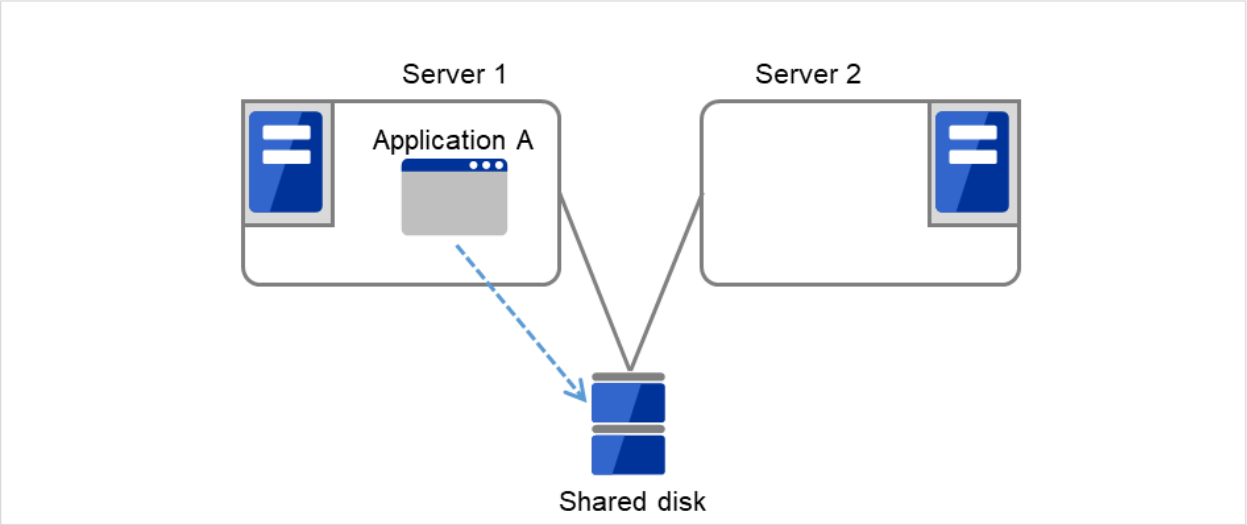


Fig. 2.3: HA cluster topology (Uni-directional standby)

A mode where there are two or more operations with each server in the cluster serving as both active and standby server, as shown in Fig. 2.4 HA cluster topology (Multi-directional standby), is referred to as multi-directional standby.

Server 1 is the active server for Application A and also the standby server for Application B.

Server 2 is the active server for Application B and also the standby server for Application A.

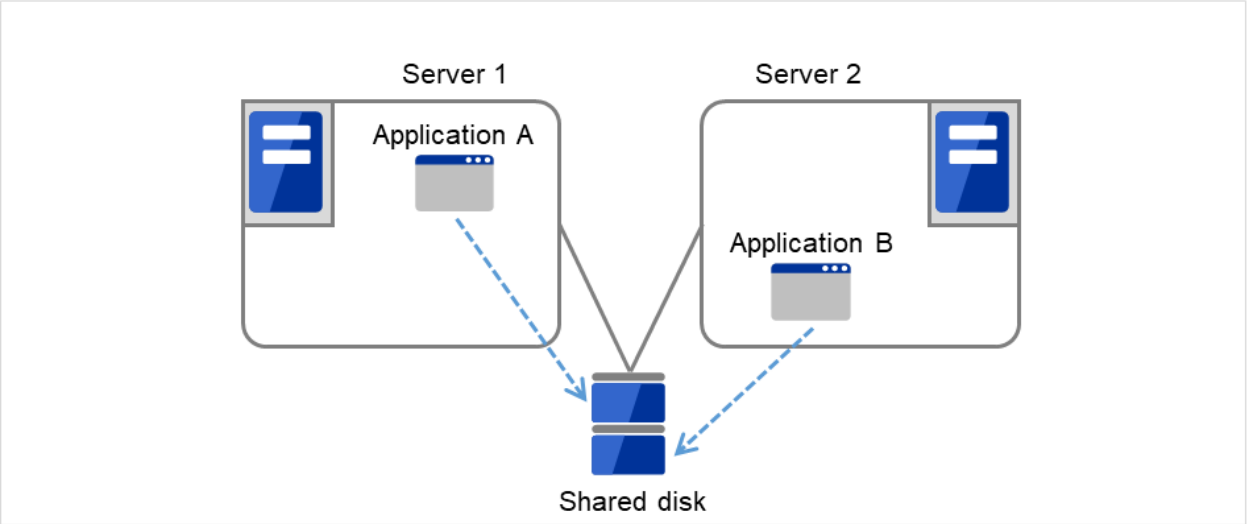


Fig. 2.4: HA cluster topology (Multi-directional standby)

2.2.2 Data mirror type

The shared disk type cluster system is good for large-scale systems. However, creating a system with this type can be costly because shared disks are generally expensive. The data mirror type cluster system provides the same functions as the shared disk type with smaller cost through mirroring of server disks.

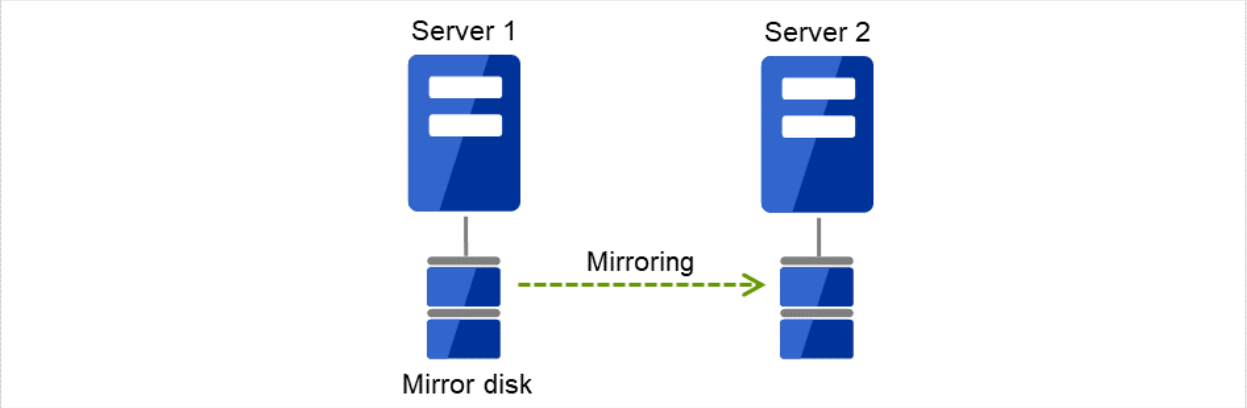


Fig. 2.5: HA cluster configuration (Data mirror type)

- Cheap since a shared disk is unnecessary.
- Ideal for the system with less data volume because of mirroring.

The data mirror type is not recommended for large-scale systems that handle a large volume of data since data needs to be mirrored between servers.

When a write request is made by an application, the data mirror engine not only writes data in the local disk but sends the write request to the standby server via the interconnect. Interconnect is a network connecting servers. It is used to monitor whether or not the server is activated in the cluster system. In addition to this purpose, interconnect is sometimes used to transfer data in the data mirror type cluster system. The data mirror engine on the standby server achieves data synchronization between standby and active servers by writing the data into the local disk of the standby server.

For read requests from an application, data is simply read from the disk on the active server.

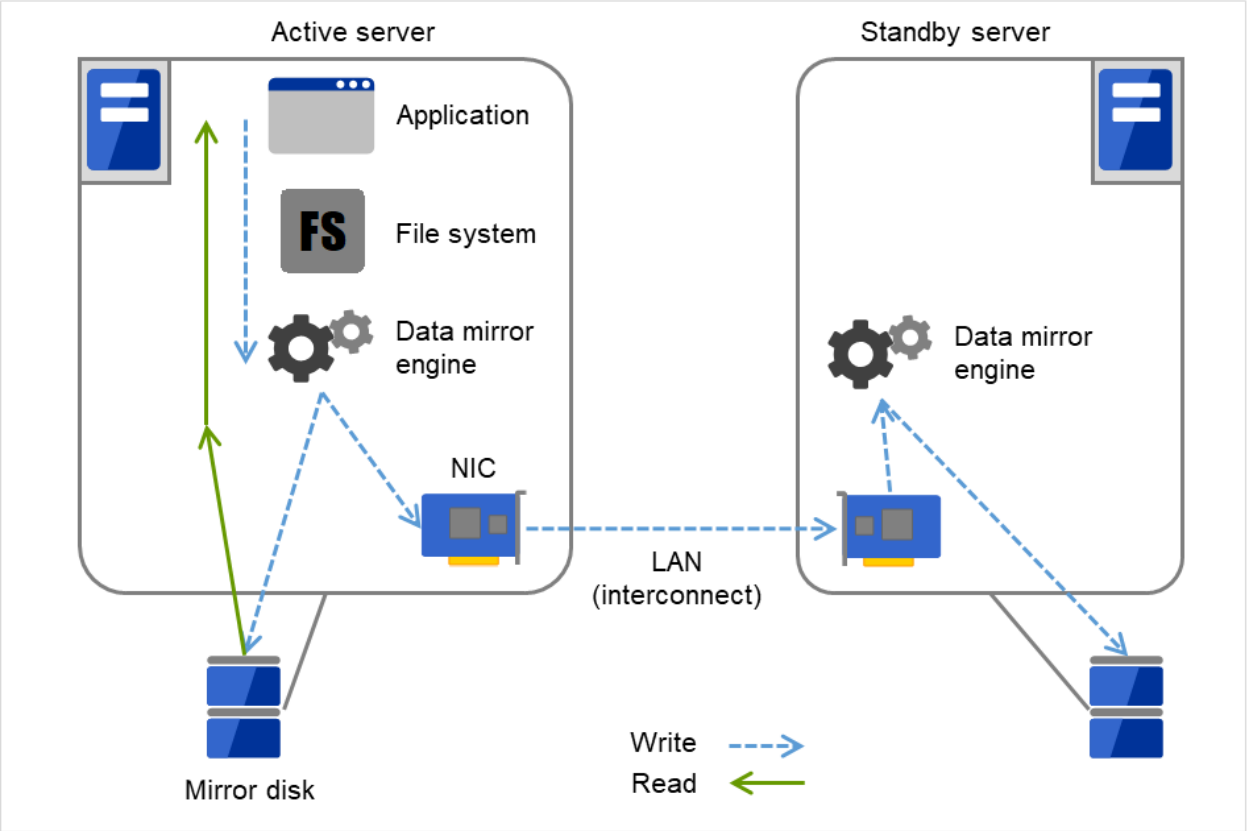


Fig. 2.6: Data mirror mechanism

Snapshot backup is applied usage of data mirroring. Because the data mirror type cluster system has shared data in two locations, you can keep the disk of the standby server as snapshot backup without spending time for backup by simply separating the server from the cluster.

Failover mechanism and its problems

There are various cluster systems such as failover clusters, load distribution clusters, and high performance computing (HPC) clusters. The failover cluster is one of the high availability (HA) cluster systems that aim to increase operational availability through establishing server redundancy and passing operations being executed to another server when a failure occurs.

2.3 Error detection mechanism

Cluster software executes failover (for example, passing operations) when a failure that can impact continued operation is detected. The following section gives you a quick view of how the cluster software detects a failure.

Heartbeat and detection of server failures

Failures that must be detected in a cluster system are failures that can cause all servers in the cluster to stop. Server failures include hardware failures such as power supply and memory failures, and OS panic. To detect such failures, heartbeat is employed to monitor whether or not the server is active.

Some cluster software programs use heartbeat not only for checking whether or not the target is active through ping response, but for sending status information on the local server. Such cluster software programs begin failover if no heartbeat response is received in heartbeat transmission, determining no response as server failure. However, grace time should be given before determining failure, since a highly loaded server can cause delay of response. Allowing grace period results in a time lag between the moment when a failure occurred and the moment when the failure is detected by the cluster software.

Detection of resource failures

Factors causing stop of operations are not limited to stop of all servers in the cluster. Failure in disks used by applications, NIC failure, and failure in applications themselves are also factors that can cause the stop of operations. These resource failures need to be detected as well to execute failover for improved availability.

Accessing a target resource is a way employed to detect resource failures if the target is a physical device. For monitoring applications, trying to service ports within the range not impacting operation is a way of detecting an error in addition to monitoring whether or not application processes are activated.

2.3.1 Problems with shared disk type

In a failover cluster system of the shared disk type, multiple servers physically share the disk device. Typically, a file system enjoys I/O performance greater than the physical disk I/O performance by keeping data caches in a server.

What if a file system is accessed by multiple servers simultaneously?

Because a general file system assumes no server other than the local updates data on the disk, inconsistency between caches and the data on the disk arises. Ultimately the data will be corrupted. The failover cluster system locks the disk device to prevent multiple servers from mounting a file system, simultaneously caused by a network partition.

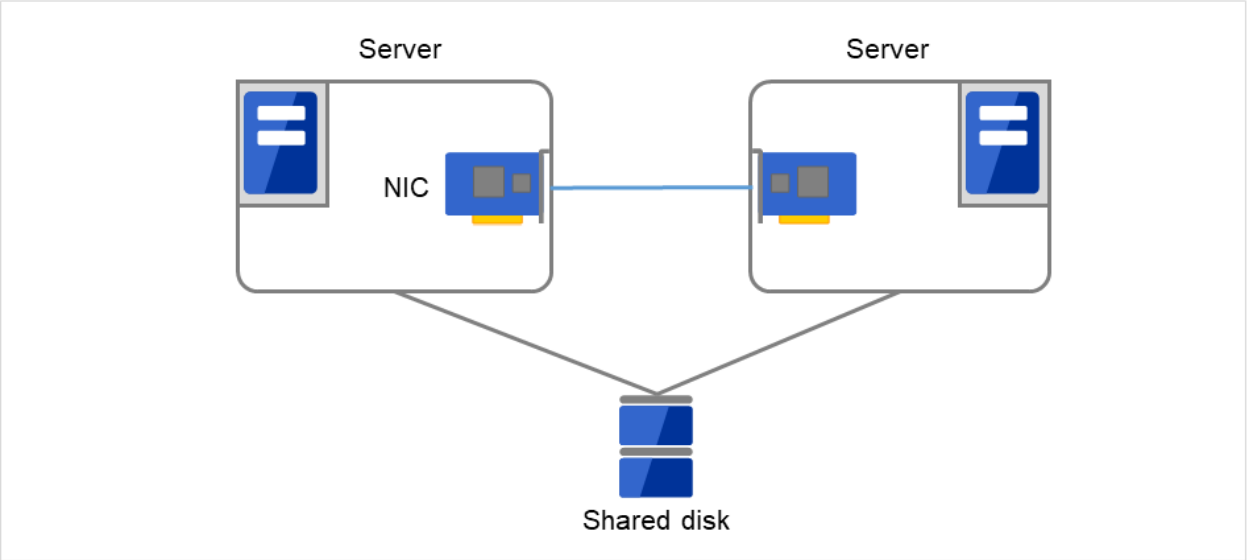


Fig. 2.7: Cluster configuration with a shared disk

2.3.2 Network partition (split-brain-syndrome)

When all interconnects between servers are disconnected, failover takes place because the servers assume other server(s) are down. To monitor whether the server is activated, a heartbeat communication is used. As a result, multiple servers mount a file system simultaneously causing data corruption. This explains the importance of appropriate failover behavior in a cluster system at the time of failure occurrence.

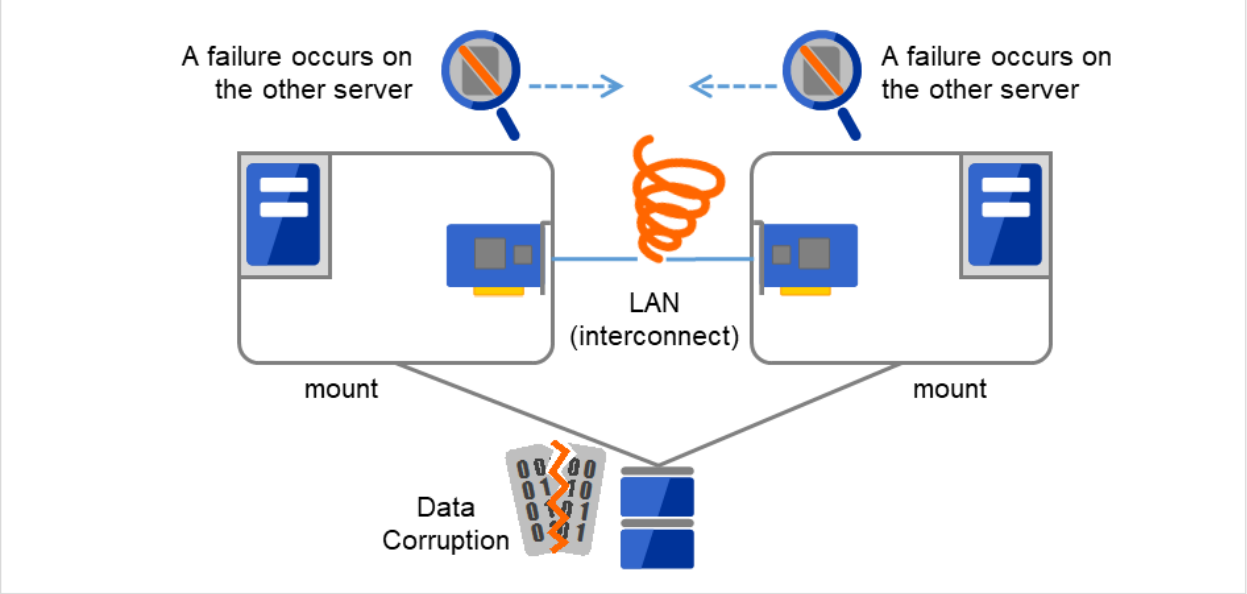


Fig. 2.8: Network partition problem

The problem explained in the section above is referred to as "network partition" or "split-brain syndrome." The failover cluster system is equipped with various mechanisms to ensure shared disk lock at the time when all interconnects are disconnected.

2.4 Taking over cluster resources

As mentioned earlier, resources to be managed by a cluster include disks, IP addresses, and applications. The functions used in the failover cluster system to inherit these resources are described below.

2.4.1 Taking over the data

Data to be passed from a server to another in a cluster system is stored in a partition on the shared disk. This means data is re-mounting the file system of files that the application uses on a healthy server. What the cluster software should do is simply mount the file system because the shared disk is physically connected to a server that inherits data.

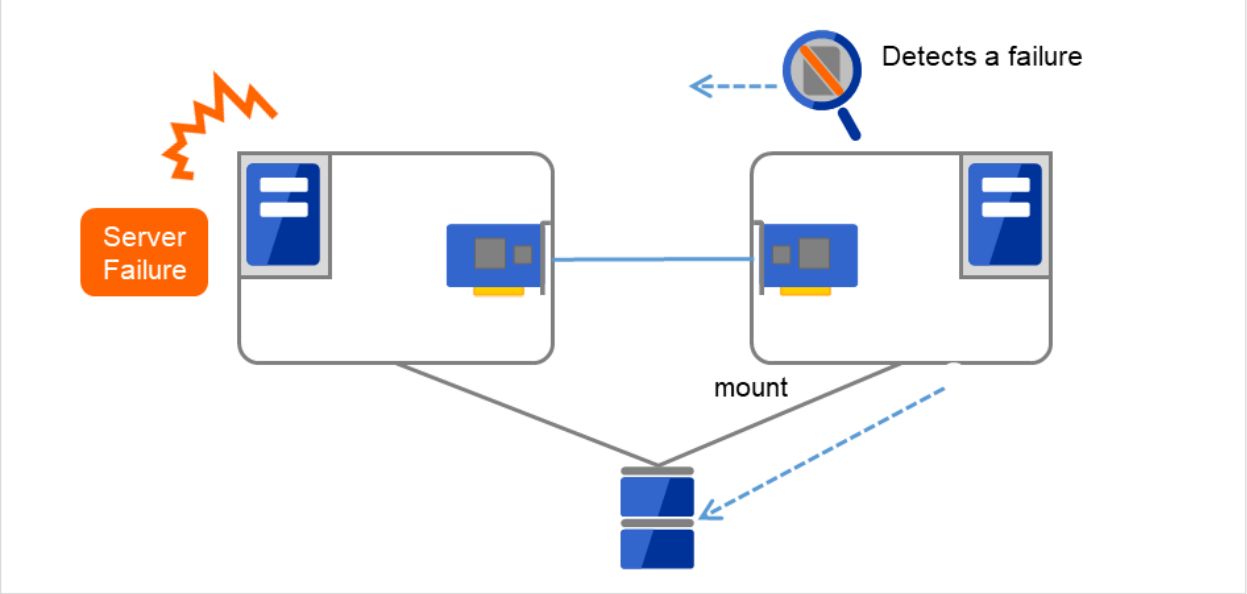


Fig. 2.9: Taking over data

"Figure 2.9 Taking over data" may look simple, but consider the following issues in designing and creating a cluster system.

One issue to consider is recovery time for a file system. A file system to be inherited may have been used by another server or being updated just before the failure occurred and requires a file system consistency check. When the file system is large, the time spent for checking consistency will be enormous. It may take a few hours to complete the check and the time is wholly added to the time for failover (time to take over operation), and this will reduce system availability.

Another issue you should consider is writing assurance. When an application writes important data into a file, it tries to ensure the data to be written into a disk by using a function such as synchronized writing. The data that the application assumes to have been written is expected to be inherited after failover. For example, a mail server reports the completion of mail receiving to other mail servers or clients after it has securely written mails it received in a spool. This will allow the spooled mail to be distributed again after the server is restarted. Likewise, a cluster system should ensure mails written into spool by a server to become readable by another server.

2.4.2 Taking over the applications

The last to come in inheritance of operation by cluster software is inheritance of applications. Unlike fault tolerant computers (FTC), no process status such as contents of memory is inherited in typical failover cluster systems. The applications running on a failed server are inherited by rerunning them on a healthy server.

For example, when instances of a database management system (DBMS) are inherited, the database is automatically recovered (roll-forward/roll-back) by startup of the instances. The time needed for this database recovery is typically a few minutes though it can be controlled by configuring the interval of DBMS checkpoint to a certain extent.

Many applications can restart operations by re-execution. Some applications, however, require going through procedures for recovery if a failure occurs. For these applications, cluster software allows to start up scripts instead of applications so that recovery process can be written. In a script, the recovery process, including cleanup of files half updated, is written as necessary according to factors for executing the script and information on the execution server.

2.4.3 Summary of failover

To summarize the behavior of cluster software:

(a) Detects a failure (heartbeat/resource monitoring)

(b) Performs fencing (resolves a network partition (NP resolution) and disconnects the failed server)

(c) Pass data

(d) Pass IP address

(e) Application Taking over

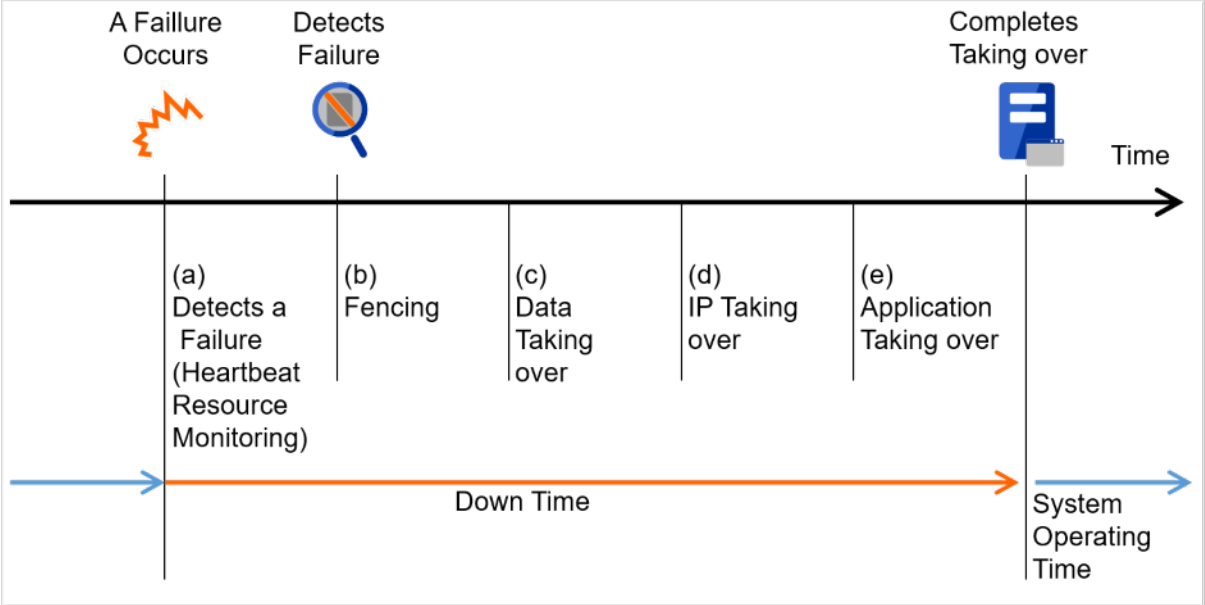


Fig. 2.10: Failover time chart

Cluster software is required to complete each task quickly and reliably (see "Figure 2.10 Failover time chart"). Cluster software achieves high availability with due consideration on what has been described so far.

2.5 Eliminating single point of failure

Having a clear picture of the availability level required or aimed is important in building a high availability system. This means when you design a system, you need to study cost effectiveness of countermeasures, such as establishing a redundant configuration to continue operations and recovering operations within a short period of time, against various failures that can disturb system operations.

Single point of failure (SPOF), as described previously, is a component where failure can lead to stop of the system. In a cluster system, you can eliminate the system's SPOF by establishing server redundancy. However, components shared among servers, such as shared disk may become a SPOF. The key in designing a high availability system is to duplicate or eliminate this shared component.

A cluster system can improve availability but failover will take a few minutes for switching systems. That means time for failover is a factor that reduces availability. Solutions for the following three, which are likely to become SPOF, will be discussed hereafter although technical issues that improve availability of a single server such as ECC memory and redundant power supply are important.

- Shared disk
- Access path to the shared disk
- LAN

2.5.1 Shared disk

Typically a shared disk uses a disk array for RAID. Because of this, the bare drive of the disk does not become SPOF. The problem is the RAID controller is incorporated. Shared disks commonly used in many cluster systems allow controller redundancy.

In general, access paths to the shared disk must be duplicated to benefit from redundant RAID controller. There are still things to be done to use redundant access paths in Linux (described later in this chapter). If the shared disk has configuration to access the same logical disk unit (LUN) from duplicated multiple controllers simultaneously, and each controller is connected to one server, you can achieve high availability by failover between nodes when an error occurs in one of the controllers.

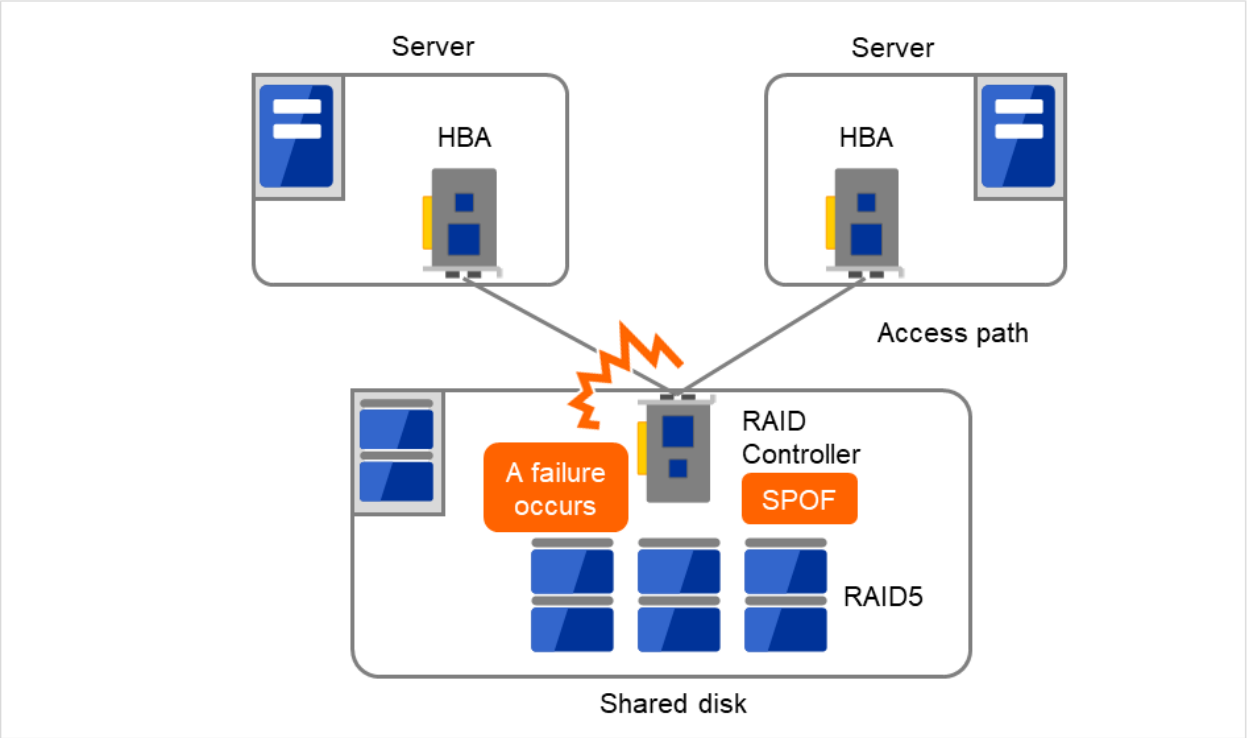


Fig. 2.11: Example of a RAID controller and access paths both being SPOF

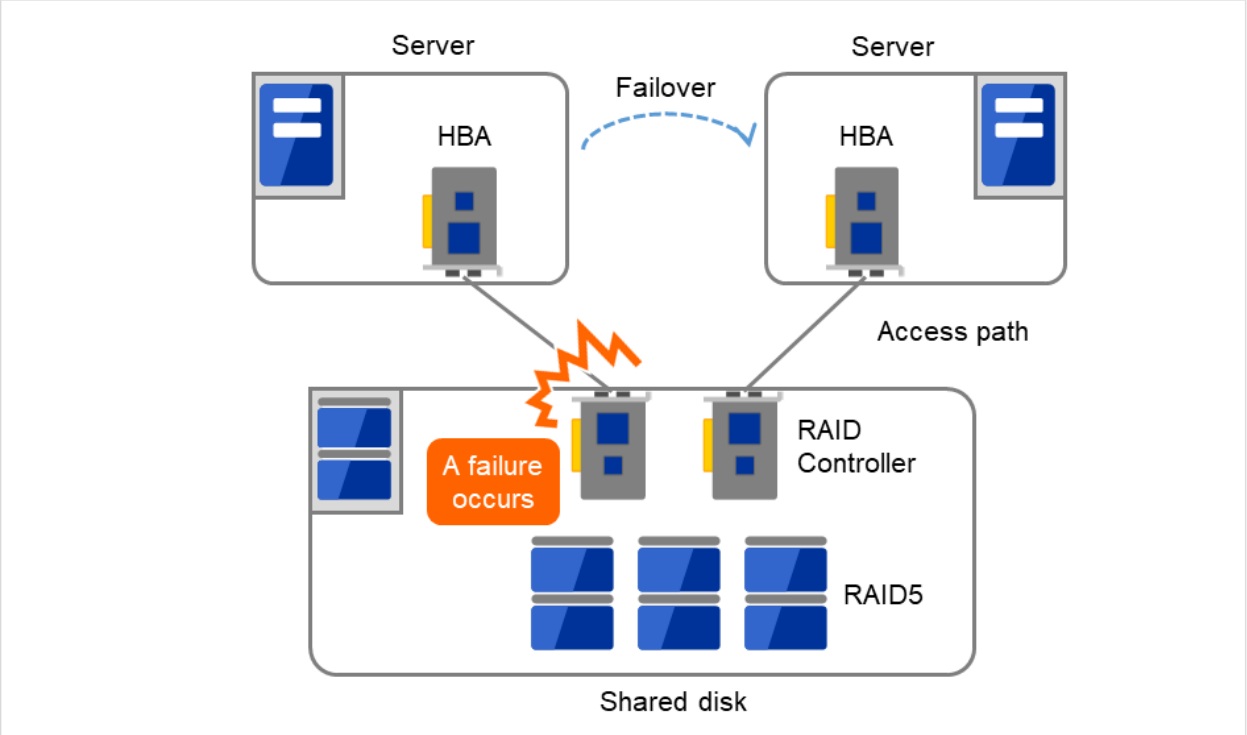


Fig. 2.12: Example of RAID controllers and access paths both being redundant

With a failover cluster system of data mirror type, where no shared disk is used, you can create an ideal system having no SPOF because all data is mirrored to the disk in the other server. However you should consider the following issues:

- Disk I/O performance in mirroring data over the network (especially writing performance)
- System performance during mirror resynchronization in recovery from server failure (mirror copy is done in the background)
- Time for mirror resynchronization (clustering cannot be done until mirror resynchronization is completed)

In a system with frequent data viewing and a relatively small volume of data, choosing the data mirror type for clustering is a key to increase availability.

2.5.2 Access path to the shared disk

In a typical configuration of the shared disk type cluster system, the access path to the shared disk is shared among servers in the cluster. To take SCSI as an example, two servers and a shared disk are connected to a single SCSI bus. A failure in the access path to the shared disk can stop the entire system.

What you can do for this is to have a redundant configuration by providing multiple access paths to the shared disk and make them look as one path for applications. The device driver allowing such is called a path failover driver. Path failover drivers are often developed and released by shared disk vendors. Path failover drivers in Linux are still under development. For the time being, as discussed earlier, offering access paths to the shared disk by connecting a server on an array controller on the shared disk basis is the way to ensure availability in Linux cluster systems.

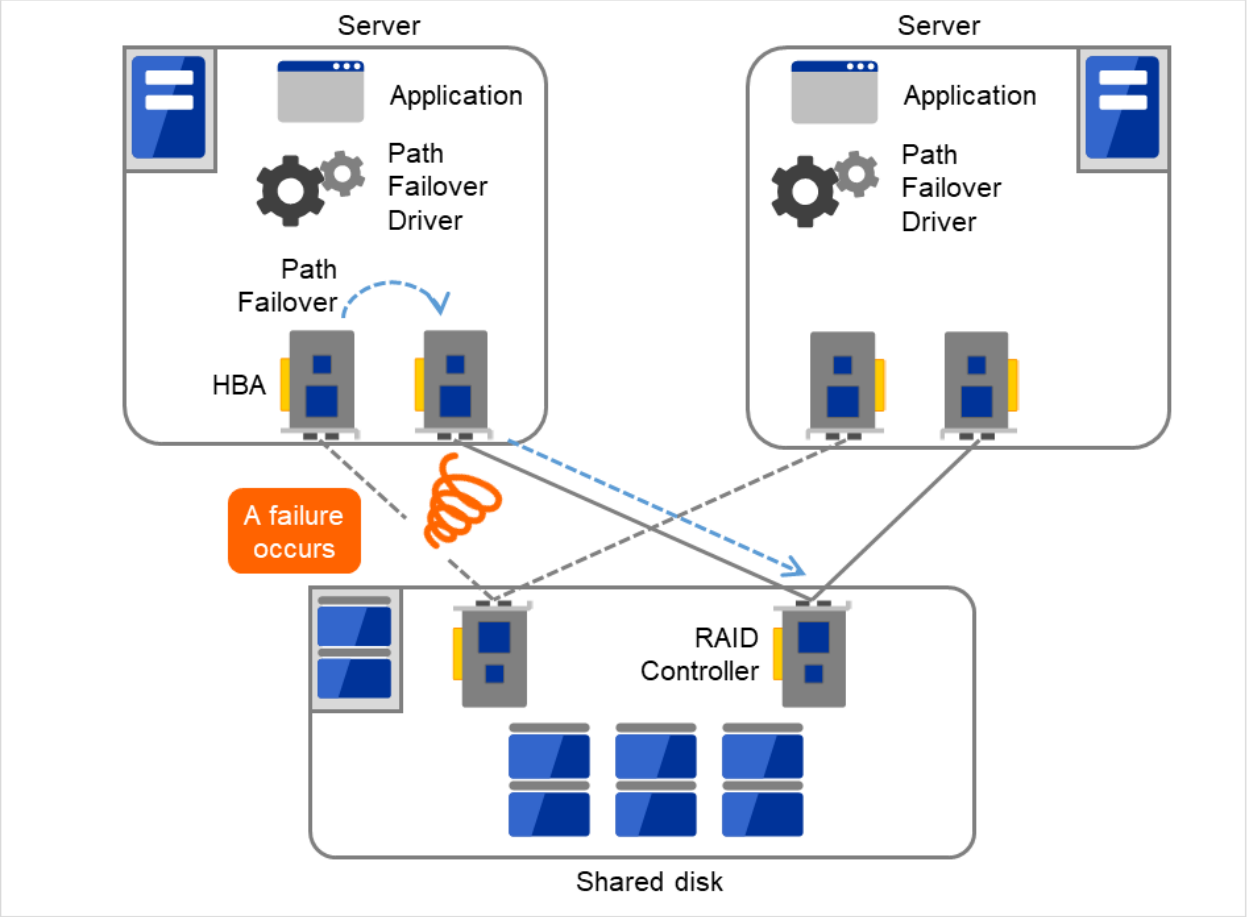


Fig. 2.13: Path failover driver

2.5.3 LAN

In any systems that run services on a network, a LAN failure is a major factor that disturbs operations of the system. If appropriate settings are made, availability of cluster system can be increased through failover between nodes at NIC failures. However, a failure in a network device that resides outside the cluster system disturbs operation of the system.

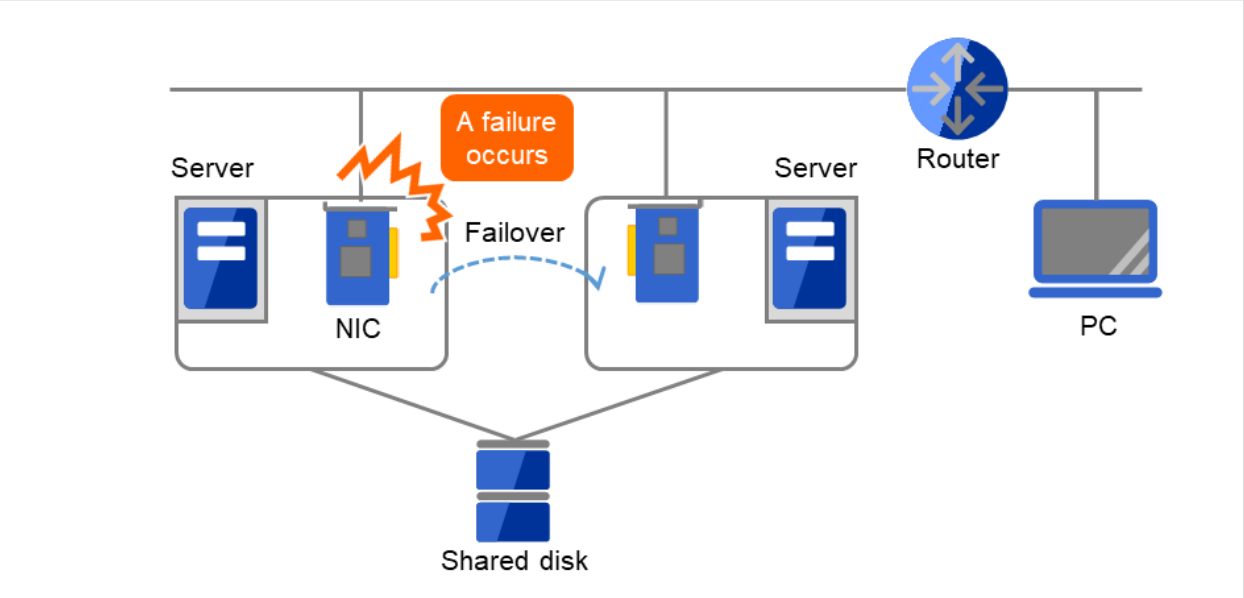


Fig. 2.14: Example of a failure with LAN (NIC)

In the case of this above figure, even if NIC on the server has a failure, a failover will keep the access from the PC to the service on the server.

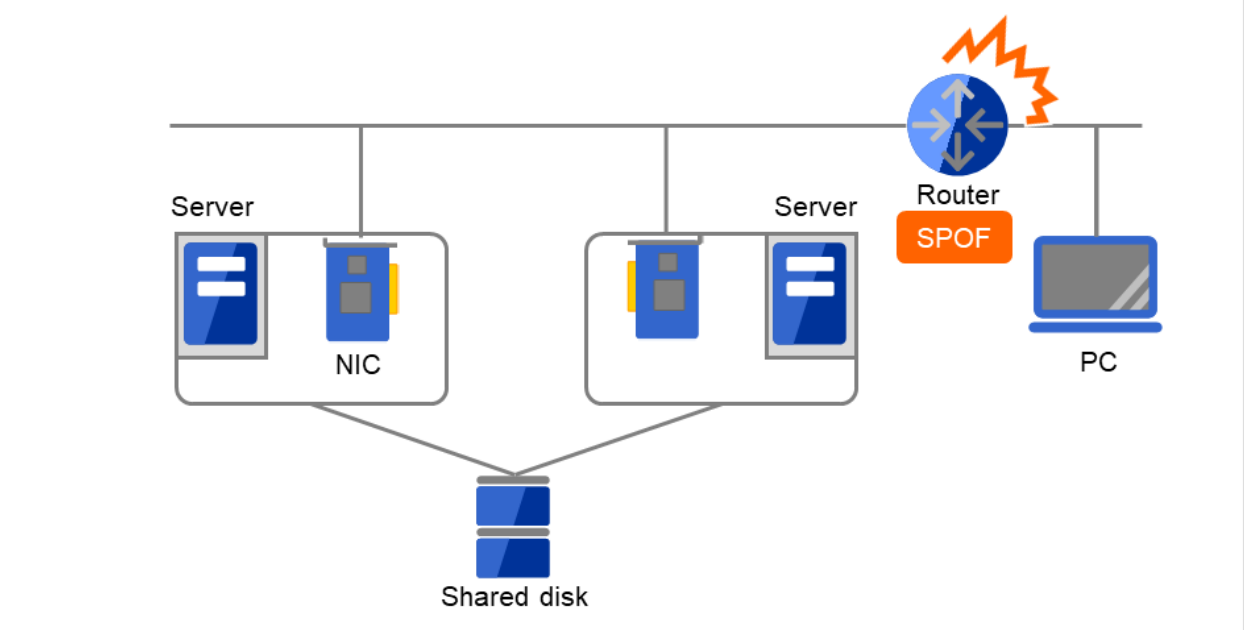


Fig. 2.15: Example of a failure with LAN (Router)

In the case of this above figure, if the router has a failure, the access from the PC to the service on the server cannot be maintained (Router becomes a SPOF).

LAN redundancy is a solution to tackle device failure outside the cluster system and to improve availability. You can apply ways used for a single server to increase LAN availability. For example, choose a primitive way to have a spare network device with its power off, and manually replace a failed device with this spare device. Choose to have a multiplex network path through a redundant configuration of high-performance network devices, and switch paths automatically. Another option is to use a driver that supports NIC redundant configuration such as Intel's ANS driver.

Load balancing appliances and firewall appliances are also network devices that are likely to become SPOF. Typically they allow failover configurations through standard or optional software. Having redundant configuration for these devices should be regarded as requisite since they play important roles in the entire system.

2.6 Operation for availability

2.6.1 Evaluation before staring operation

Given many of factors causing system troubles are said to be the product of incorrect settings or poor maintenance, evaluation before actual operation is important to realize a high availability system and its stabilized operation. Exercising the following for actual operation of the system is a key in improving availability:

- Clarify and list failures, study actions to be taken against them, and verify effectiveness of the actions by creating dummy failures.
- Conduct an evaluation according to the cluster life cycle and verify performance (such as at degenerated mode)
- Arrange a guide for system operation and troubleshooting based on the evaluation mentioned above.

Having a simple design for a cluster system contributes to simplifying verification and improvement of system availability.

2.6.2 Failure monitoring

Despite the above efforts, failures still occur. If you use the system for long time, you cannot escape from failures: hardware suffers from aging deterioration and software produces failures and errors through memory leaks or operation beyond the originally intended capacity. Improving availability of hardware and software is important yet monitoring for failure and troubleshooting problems is more important. For example, in a cluster system, you can continue running the system by spending a few minutes for switching even if a server fails. However, if you leave the failed server as it is, the system no longer has redundancy and the cluster system becomes meaningless should the next failure occur.

If a failure occurs, the system administrator must immediately take actions such as removing a newly emerged SPOF to prevent another failure. Functions for remote maintenance and reporting failures are very important in supporting services for system administration. Linux is known for providing good remote maintenance functions. Mechanism for reporting failures are coming in place. To achieve high availability with a cluster system, you should:

- Remove or have complete control on single point of failure.
- Have a simple design that has tolerance and resistance for failures, and be equipped with a guide for operation and troubleshooting.
- Detect a failure quickly and take appropriate action against it.

CHAPTER

THREE

USING EXPRESSCLUSTER

This chapter explains the components of EXPRESSCLUSTER, how to design a cluster system, and how to use EXPRESSCLUSTER.

This chapter covers:

3.1 What is EXPRESSCLUSTER?

EXPRESSCLUSTER is software that enhances availability and expandability of systems by a redundant (clustered) system configuration. The application services running on the active server are automatically inherited to a standby server when an error occurs in the active server.

3.2 EXPRESSCLUSTER modules

EXPRESSCLUSTER consists of following two modules:

- **EXPRESSCLUSTER Server**
  A core component of EXPRESSCLUSTER. This includes all high availability functions of the server. The server functions of the Cluster WebUI, are also included.
- **Cluster WebUI**
  This is a tool to create the configuration data of EXPRESSCLUSTER and to manage EXPRESSCLUSTER operations. Uses a Web browser as a user interface. The Cluster WebUI is installed in EXPRESSCLUSTER Server, but it is distinguished from the EXPRESSCLUSTER Server because the Cluster WebUI is operated from the Web browser on the management PC.

3.3 Software configuration of EXPRESSCLUSTER

The software configuration of EXPRESSCLUSTER should look similar to the figure below. Install the EXPRESSCLUSTER Server (software) on a Linux server, and the Cluster WebUI on a management PC or a server. Because the main functions of Cluster WebUI are included in EXPRESSCLUSTER Server, it is not necessary to separately install them. The Cluster WebUI can be used through the web browser on the management PC or on each server in the cluster.

(a) EXPRESSCLUSTER Server

(b) Cluster WebUI

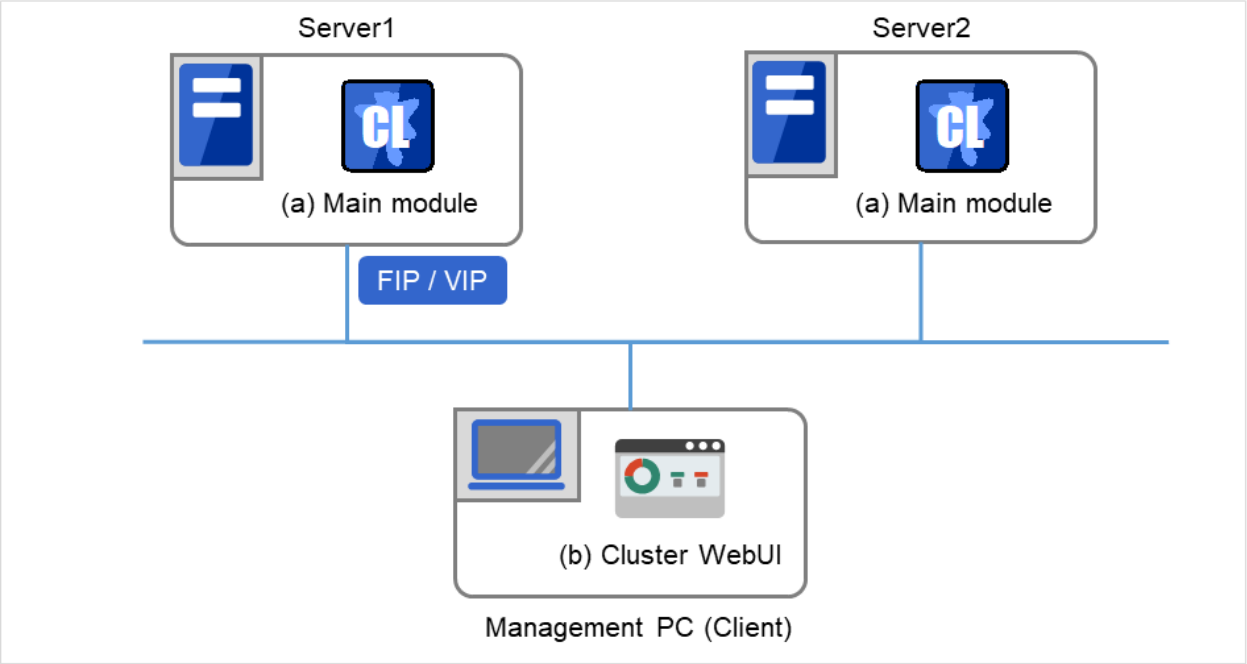


Fig. 3.1: Software configuration of EXPRESSCLUSTER

3.3.1 How an error is detected in EXPRESSCLUSTER

There are three kinds of monitoring in EXPRESSCLUSTER: (1) server monitoring, (2) application monitoring, and (3) internal monitoring. These monitoring functions let you detect an error quickly and reliably. The details of the monitoring functions are described below.

3.3.2 What is server monitoring?

Server monitoring is the most basic function of the failover-type cluster system. It monitors if a server that constitutes a cluster is properly working.
EXPRESSCLUSTER regularly checks whether other servers are properly working in the cluster system. This way of verification is called "heartbeat communication." The heartbeat communication uses the following communication paths:

- **Primary Interconnect**

Uses an Ethernet NIC in communication path dedicated to the failover-type cluster system. This is used to exchange information between the servers as well as to perform heartbeat communication.

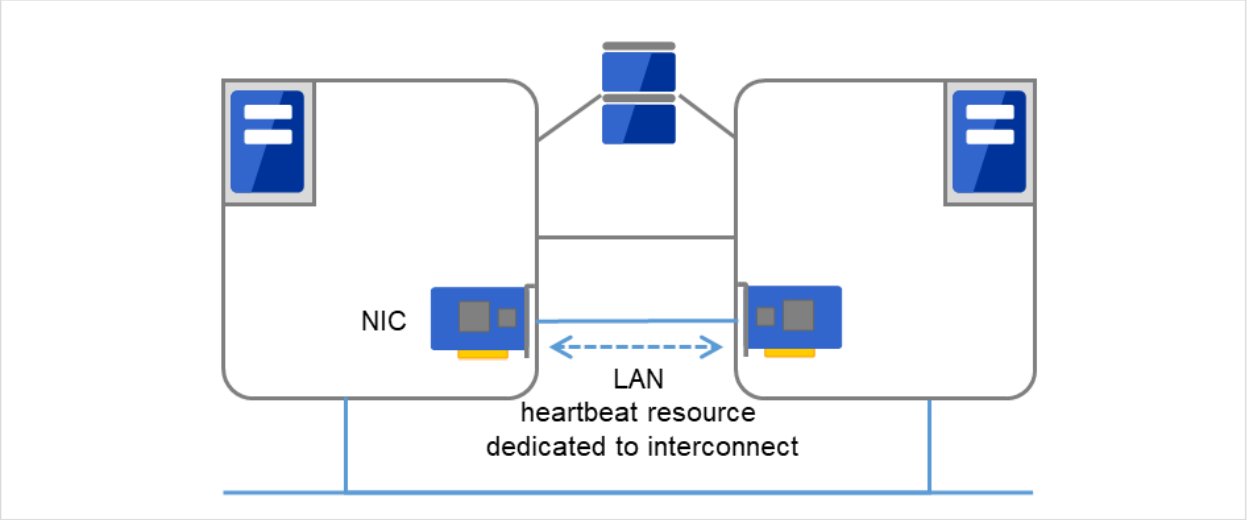


Fig. 3.2: LAN heartbeat/Kernel mode LAN heartbeat (Primary interconnect)

- **Secondary Interconnect**
  Uses a communication path used for communication with client machine as an alternative interconnect. Any Ethernet NIC can be used as long as TCP/IP can be used. This is also used to exchange information between the servers and to perform heartbeat communication.

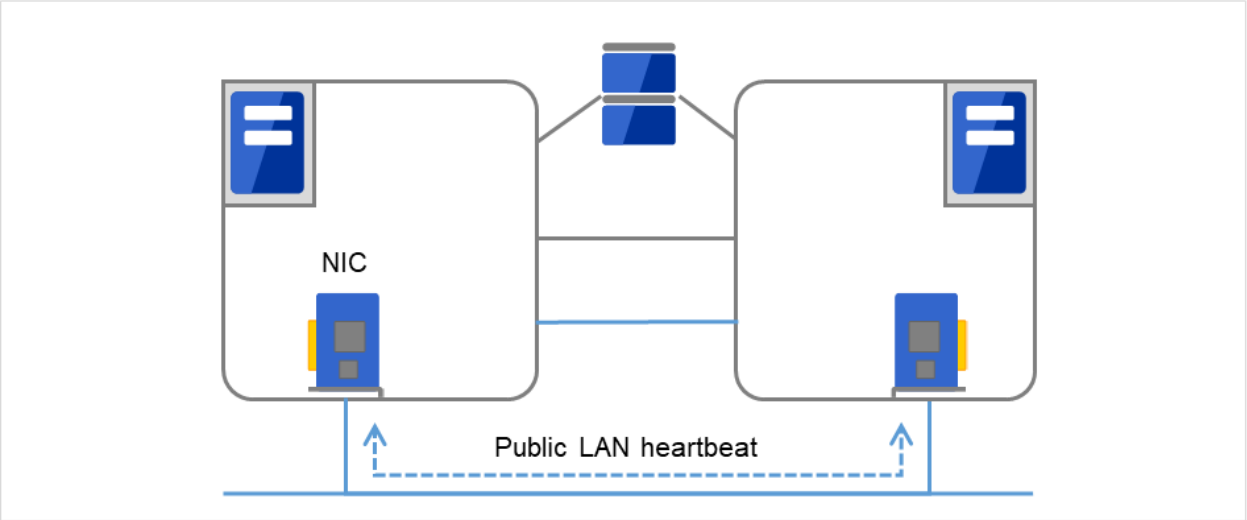


Fig. 3.3: LAN heartbeat/Kernel mode LAN heartbeat (Secondary interconnect)

- **Shared disk**
  Creates an EXPRESSCLUSTER-dedicated partition (EXPRESSCLUSTER partition) on the disk that is connected to all servers that constitute the failover-type cluster system, and performs heartbeat communication on the EXPRESSCLUSTER partition.

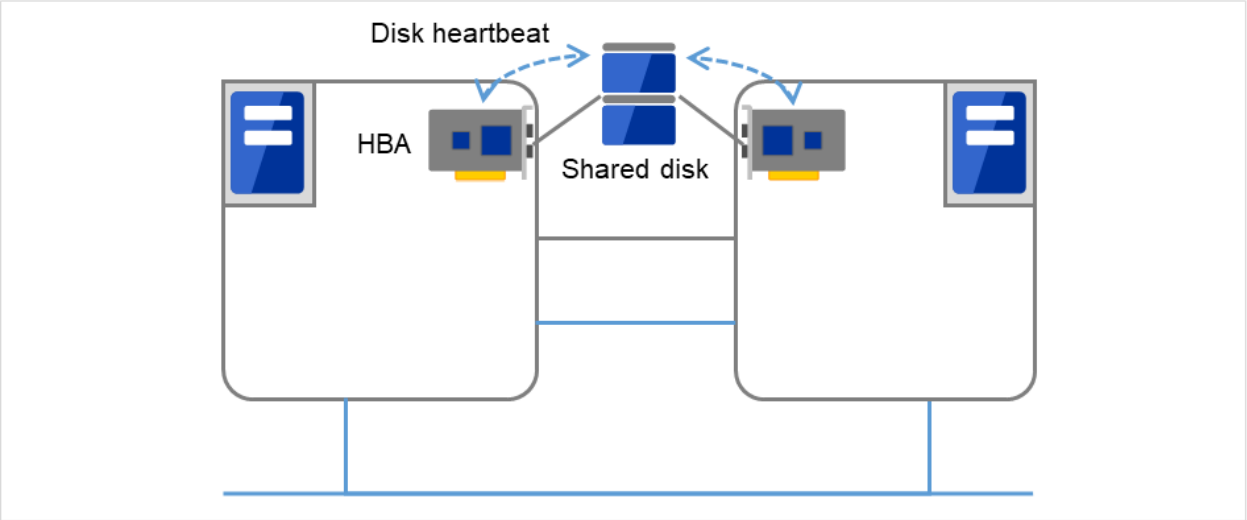


Fig. 3.4: Disk heartbeat

- **Witness**
  This is used by the external Witness server running the Witness server service to check if other servers constructing the failover-type cluster exist through communication with them.

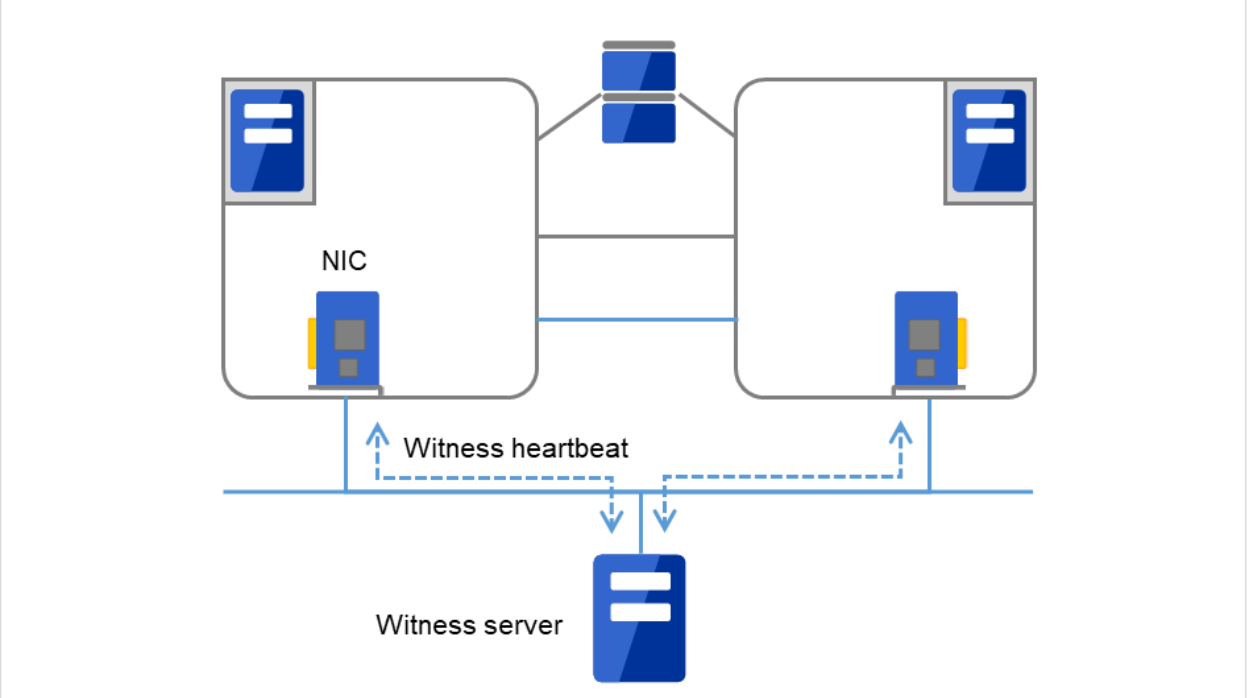


Fig. 3.5: Witness heartbeat

Having these communication paths dramatically improves the reliability of the communication between the servers, and prevents the occurrence of network partition.

Note: Network partition refers to a condition when a network gets split by having a problem in all communication

paths of the servers in a cluster. In a cluster system that is not capable of handling a network partition, a problem occurred in a communication path and a server cannot be distinguished. As a result, multiple servers may access the same resource and cause the data in a cluster system to be corrupted.

3.3.3 What is application monitoring?

Application monitoring is a function that monitors applications and factors that cause a situation where an application cannot run.

- **Activation status of application monitoring**
  An error can be detected by starting up an application from an exec resource in EXPRESSCLUSTER and regularly checking whether a process is active or not by using the pid monitor resource. It is effective when the factor for application to stop is due to error termination of an application.

  Note:

  - An error in resident process cannot be detected in an application started up by EXPRESSCLUSTER. When the monitoring target application starts and stops a resident process, an internal application error (such as application stalling, result error) cannot be detected.

- **Resource monitoring**
  An error can be detected by monitoring the cluster resources (such as disk partition and IP address) and public LAN using the monitor resources of the EXPRESSCLUSTER. It is effective when the factor for application to stop is due to an error of a resource which is necessary for an application to operate.

3.3.4 What is internal monitoring?

Internal monitoring refers to an inter-monitoring of modules within EXPRESSCLUSTER. It monitors whether each monitoring function of EXPRESSCLUSTER is properly working. Activation status of EXPRESSCLUSTER process monitoring is performed within EXPRESSCLUSTER.

- Critical monitoring of EXPRESSCLUSTER process

3.3.5 Monitorable and non-monitorable errors

There are monitorable and non-monitorable errors in EXPRESSCLUSTER. It is important to know what can or cannot be monitored when building and operating a cluster system.

3.3.6 Detectable and non-detectable errors by server monitoring

Monitoring condition: A heartbeat from a server with an error is stopped

- Example of errors that can be monitored:
  - Hardware failure (of which OS cannot continue operating)
  - System panic
- Example of error that cannot be monitored:
  - Partial failure on OS (for example, only a mouse or keyboard does not function)

3.3.7 Detectable and non-detectable errors by application monitoring

Monitoring conditions: Termination of applications with errors, continuous resource errors, and disconnection of a path to the network devices.

- Example of errors that can be monitored:
  - Abnormal termination of an application
  - Failure to access the shared disk (such as HBA[1] failure)
  - Public LAN NIC problem
- Example of errors that cannot be monitored:
  - Application stalling and resulting in error. EXPRESSCLUSTER cannot monitor application stalling and error results. However, it is possible to perform failover by creating a program that monitors applications and terminates itself when an error is detected, starting the program using the exec resource, and monitoring application using the PID monitor resource.

[1] HBA is an abbreviation for host bus adapter. This adapter is not for the shared disk, but for the server.

3.4 Fencing Function

EXPRESSCLUSTER's fencing function consists of network partition resolution and forced stopping.

3.4.1 Network partition resolution

Upon detecting that a heartbeat from a server is interrupted, EXPRESSCLUSTER determines whether the cause of this interruption is an error in a server or a network partition. If it is judged as a server failure, failover (activate resources and start applications on a healthy server) is performed. If it is judged as a network partition, protecting data is given priority over operations and a processing such as emergency shutdown is performed.
The following is the network partition resolution method:

- ping method
- http method

See also:

For the details on the network partition resolution method, see "Details on network partition resolution resources" of the Reference Guide.

3.4.2 Forced stop

When a server failure is detected, a healthy server can send a stop request to the failed server. Making the failed server stop eliminates the possibility of simultaneously starting business applications on two or more servers. The forced stop is made before a failover is started.

See also:

For the details on the forced stop function, see "Forced stop resource details" in the "Reference Guide".

3.5 Failover mechanism

Upon detecting that a heartbeat from a server is interrupted, EXPRESSCLUSTER determines whether the cause of this interruption is an error in a server or a network partition before starting a failover. Then a failover is performed by activating various resources and starting up applications on a properly working server.

The group of resources which fail over at the same time is called a "failover group." From a user's point of view, a failover group appears as a virtual computer.

Note: In a cluster system, a failover is performed by restarting the application from a properly working node. Therefore, what is saved in an application memory cannot be failed over.

From occurrence of error to completion of failover takes a few minutes. See the "Figure 3.6 Failover time chart" below:

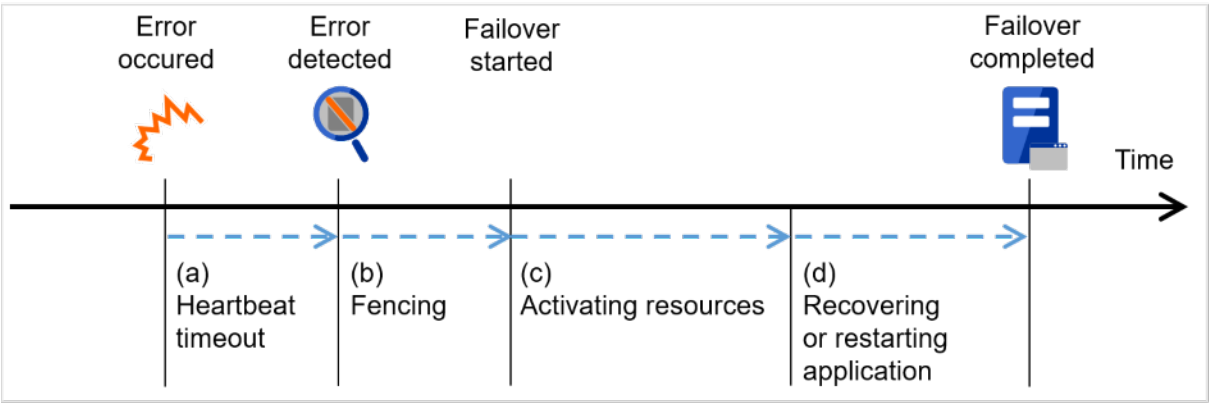


Fig. 3.6: Failover time chart

a. **Heartbeat timeout**
   - The time for a standby server to detect an error after that error occurred on the active server.
   - The setting values of the cluster properties should be adjusted depending on the application load. (The default value is 90 seconds.)

b. **Fencing**
   - The time for network partition resolution and forced stopping.
   - For network partition resolution, EXPRESSCLUSTER checks whether stop of heartbeat (heartbeat timeout) detected from the other server is due to a network partition or an error in the other server.
     Confirmation completes immediately.
   - For forced stopping, a stop request is sent to the server that is recognized to be the failure source.
     How long it will take varies depending on the cluster's operating environment such as a physical one, a virtual one, or the cloud.

c. **Activating various resources**
   - The time to activate the resources necessary for operating an application.
   - The file system recovery, transfer of data in disks, and transfer of IP addresses are performed.

- The resources can be activated in a few seconds in ordinary settings, but the required time changes depending on the type and the number of resources registered to the failover group. For more information, refer to the "Installation and Configuration Guide".

d. **Start script execution time**

- The data recovery time for a roll-back or roll-forward of the database and the startup time of the application to be used in operation.
- The time for roll-back or roll-forward can be predicted by adjusting the check point interval. For more information, refer to the document that comes with each software product.

3.5.1 Failover resources

EXPRESSCLUSTER can fail over the following resources:

- **Switchable partition**
  - Resources such as disk resource, mirror disk resource and hybrid disk resource.
  - A disk partition to store the data that the application takes over.
- **Floating IP Address**
  - By connecting an application using the floating IP address, a client does not have to be conscious about switching the servers due to failover processing.
  - It is achieved by dynamic IP address allocation to the public LAN adapter and sending ARP packet. Connection by floating IP address is possible from most of the network devices.
- **Script (exec resource)**
  - In EXPRESSCLUSTER, applications are started up from the scripts.
  - The file failed over on the shared disk may not be complete as data even if it is properly working as a file system. Write the recovery processing specific to an application at the time of failover in addition to the startup of an application in the scripts.

Note: In a cluster system, failover is performed by restarting the application from a properly working node. Therefore, what is saved in an application memory cannot be failed over.

3.5.2 System configuration of the failover-type cluster

In a failover-type cluster, a disk array device is shared between the servers in a cluster. When an error occurs on a server, the standby server takes over the applications using the data on the shared disk.

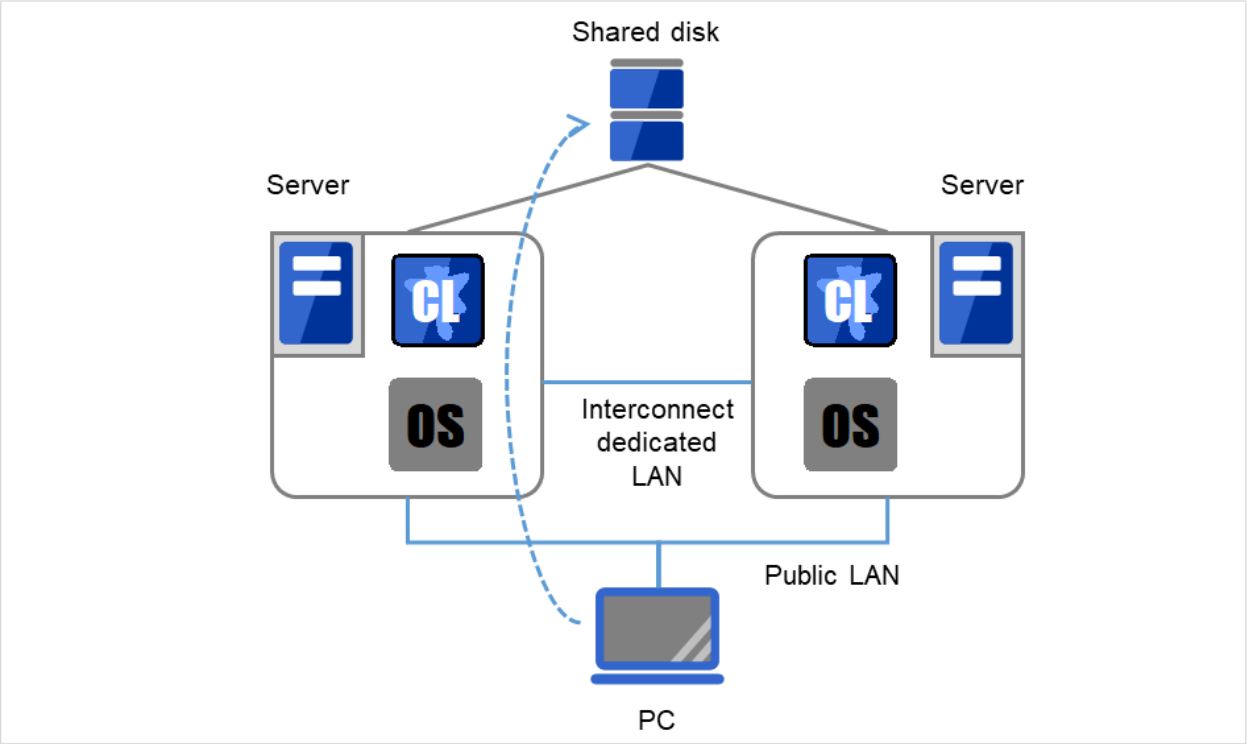


Fig. 3.7: System configuration of failover-type cluster

A failover-type cluster can be divided into the following categories depending on the cluster topologies:

Uni-Directional Standby Cluster System

In the uni-directional standby cluster system, the active server runs applications while the other server, the standby server, does not. This is the simplest cluster topology and you can build a high-availability system without performance degradation after failing over.

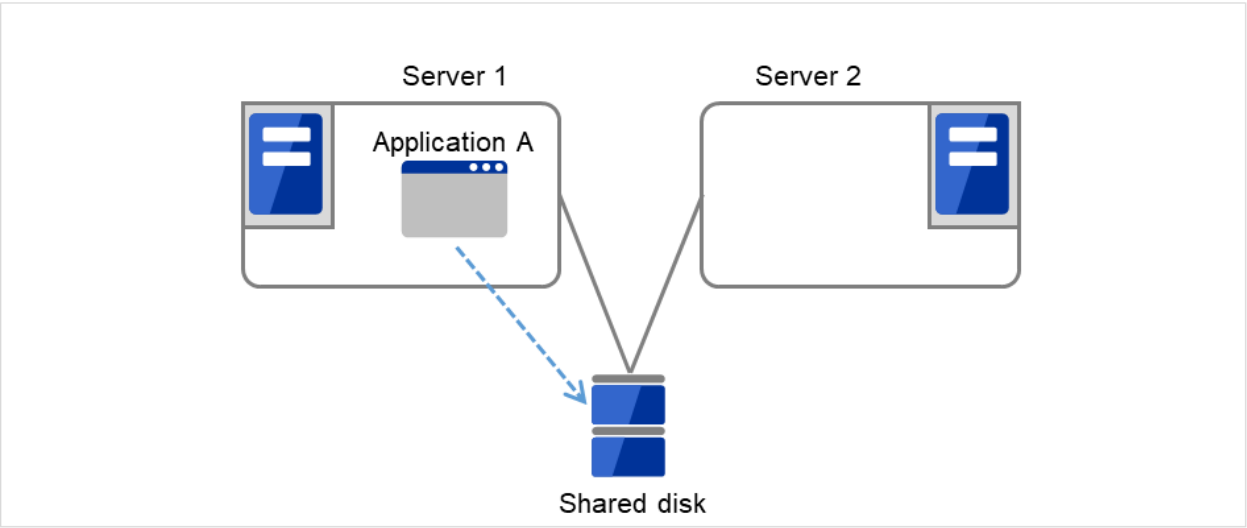


Fig. 3.8: Uni-directional standby cluster (1)

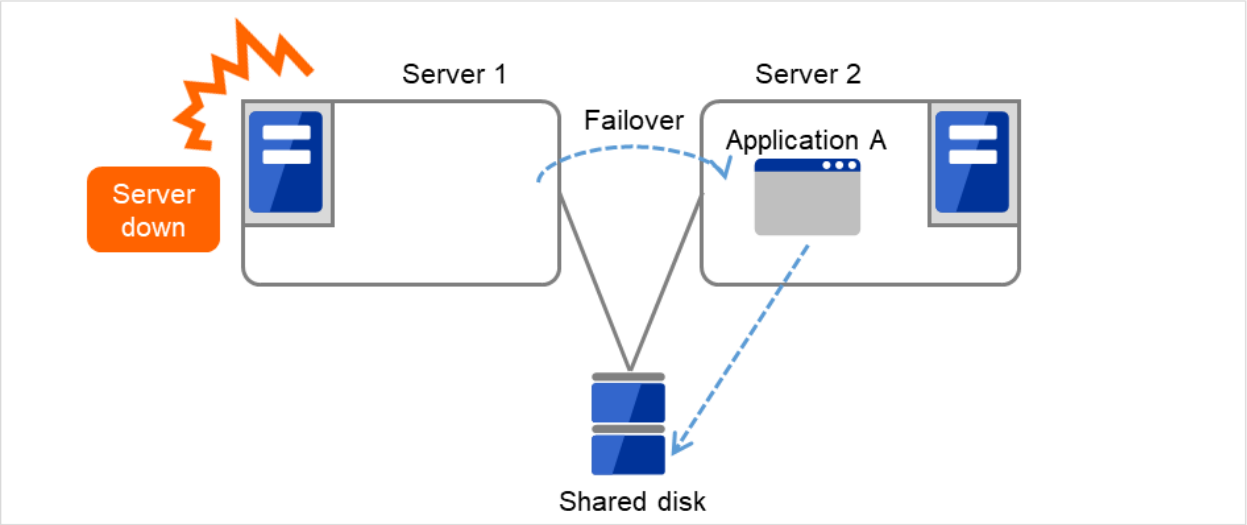


Fig. 3.9: Uni-directional standby cluster (2)

Multi-directional standby cluster system with the same application

In the same application multi-directional standby cluster system, the same applications are activated on multiple servers. These servers also operate as standby servers. The applications must support multi-directional standby operation. When the application data can be split into multiple data, depending on the data to be accessed, you can build a load distribution system per data partitioning basis by changing the client's connecting server.

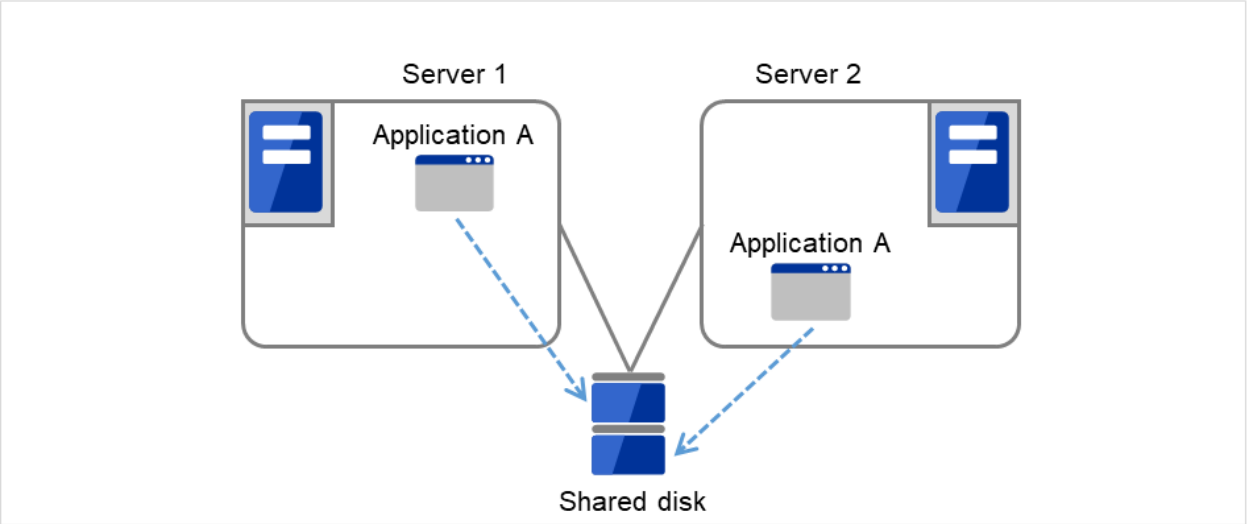


Fig. 3.10: Multi-directional standby cluster system with the same application (1)

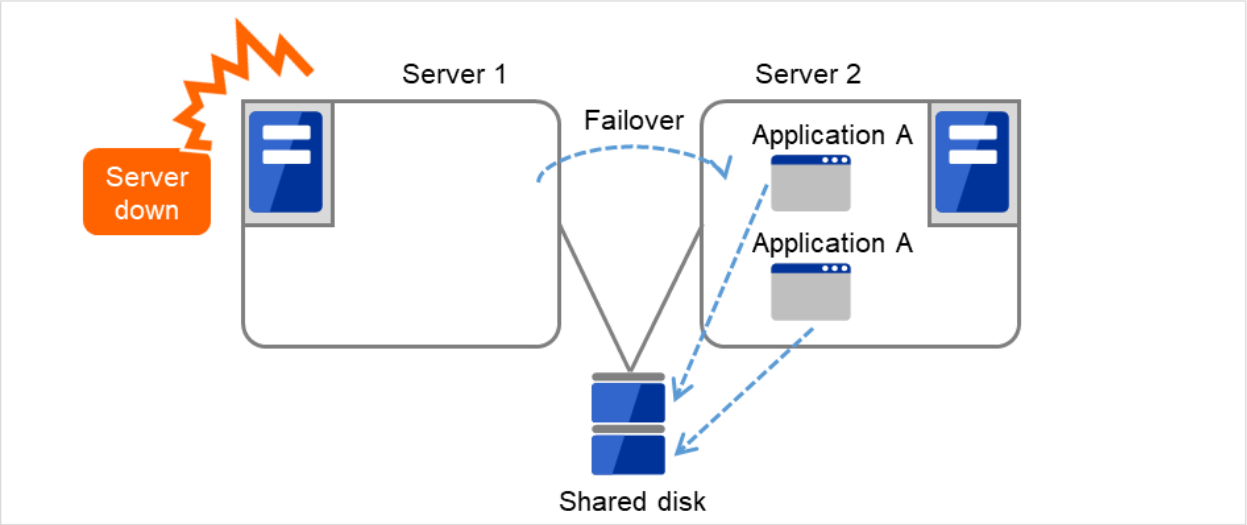


Fig. 3.11: Multi-directional standby cluster system with the same application (2)

Multi-directional standby cluster system with different applications

In the different application multi-directional standby cluster system, different applications are activated on multiple servers and these servers also operate as standby servers. The applications do not have to support multi-directional standby operation. A load distribution system can be built per application unit basis.

Application A and Application B are different applications.

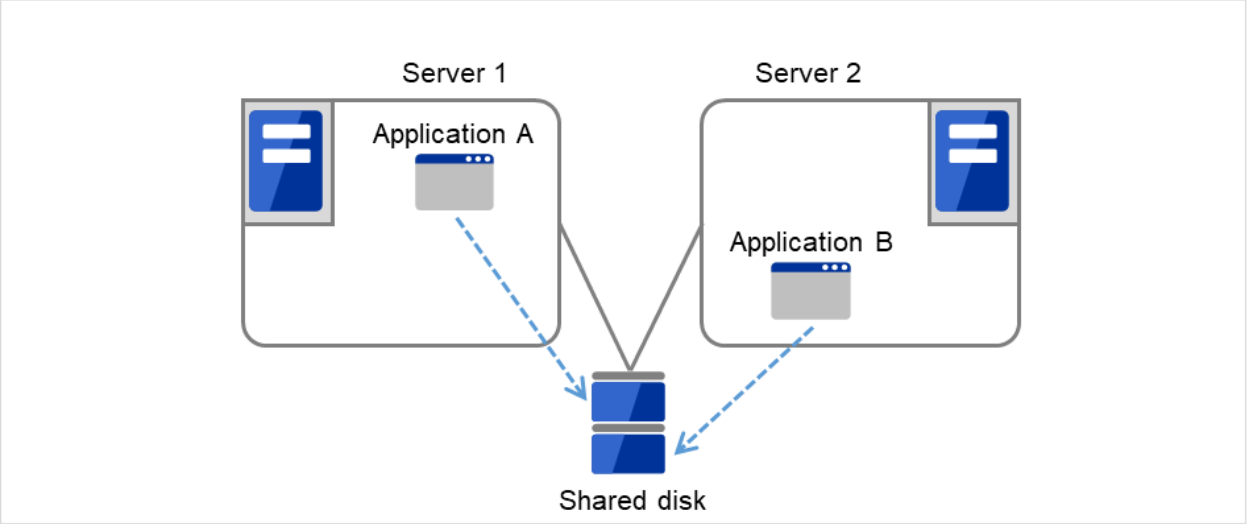


Fig. 3.12: Multi-directional standby cluster system with different applications (1)

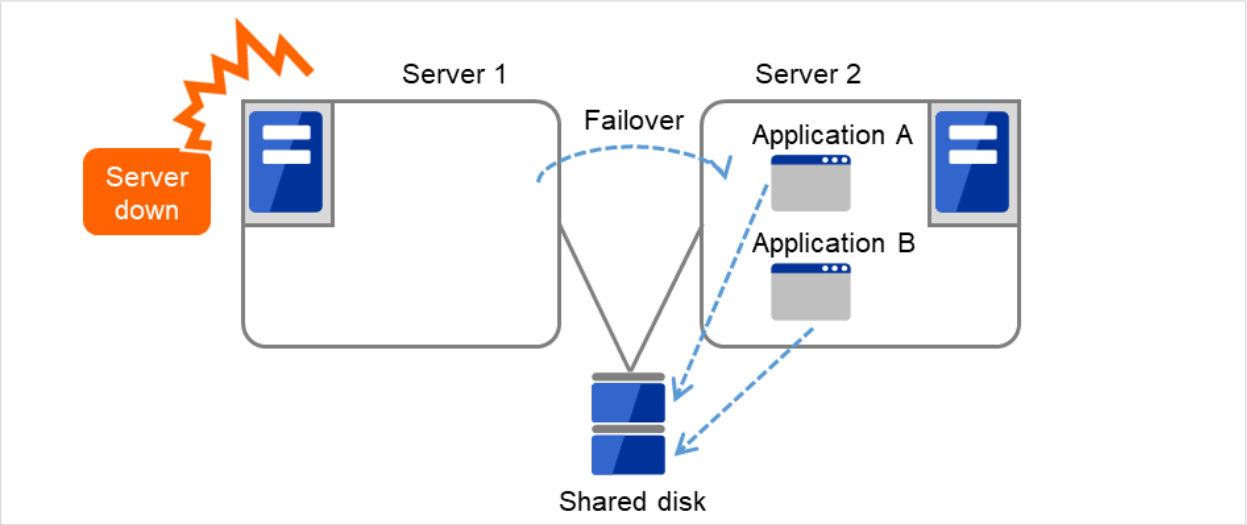


Fig. 3.13: Multi-directional standby cluster system with different applications (2)

Node to Node Configuration

The configuration can be expanded with more nodes by applying the configurations introduced thus far. In a node to node configuration described below, three different applications are run on three servers and one standby server takes over the application if any problem occurs. In a uni-directional standby cluster system, one of the two servers functions as a standby server. However, in a node to node configuration, only one of the four server functions as a standby server and performance deterioration is not anticipated if an error occurs only on one server.

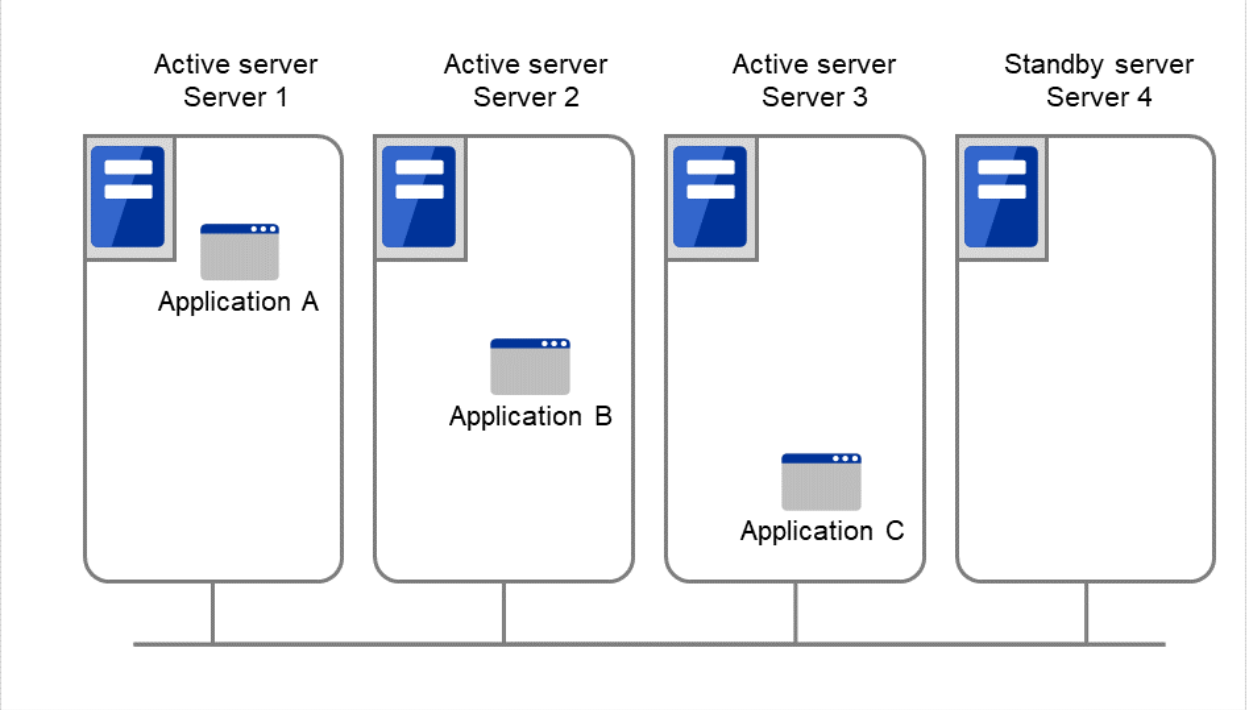


Fig. 3.14: Node to node configuration (1)

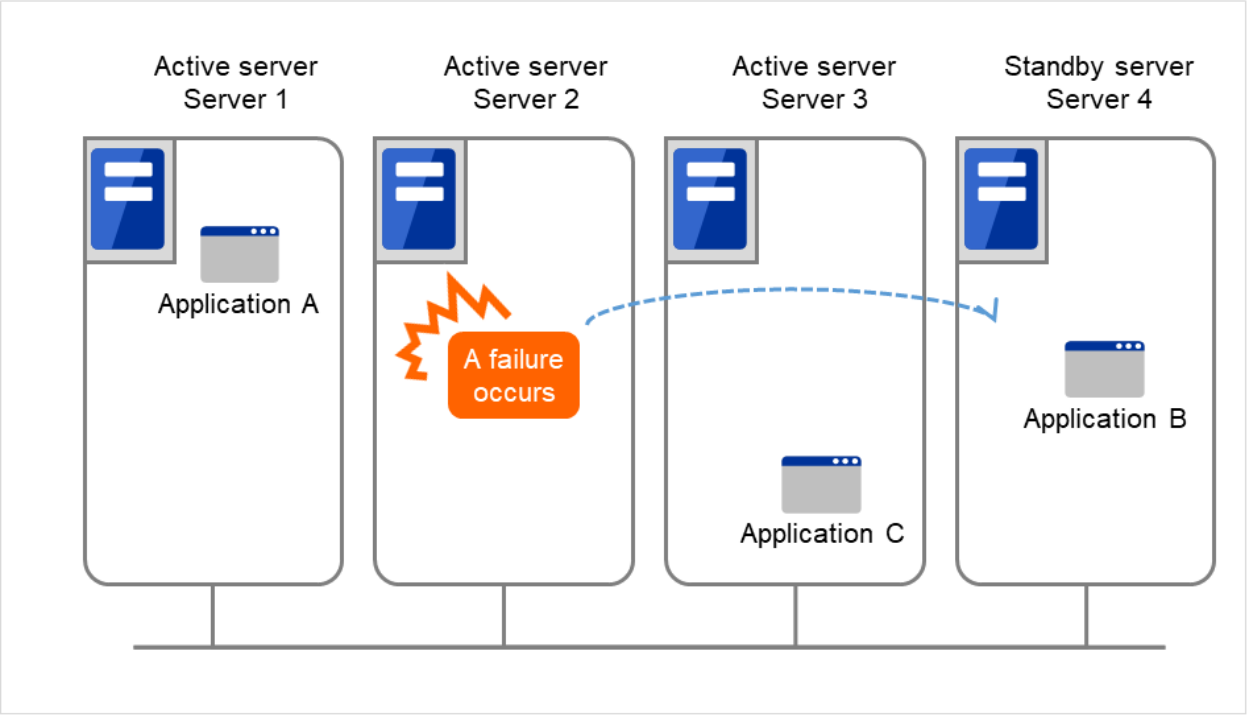


Fig. 3.15: Node to node configuration (2)

3.5.3 Hardware configuration of the shared disk type cluster

The hardware configuration of the shared disk in EXPRESSCLUSTER is described below. In general, the following is used for communication between the servers in a cluster system:

- Two NIC cards (one for external communication, one for EXPRESSCLUSTER)
- Specific space of a shared disk

SCSI or FibreChannel can be used for communication interface to a shared disk; however, recently FibreChannel is more commonly used.

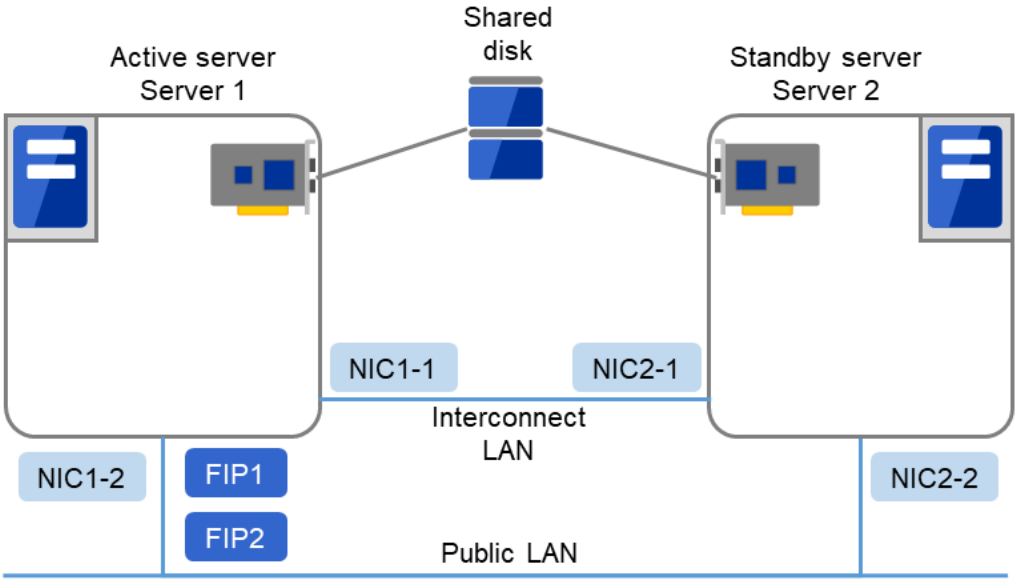


Fig. 3.16: Example of cluster configuration (Shared disk type)

FIP1	10.0.0.11 (Access destination from the Cluster WebUI client)
FIP2	10.0.0.12 (Access destination from the operation client)
NIC1-1	192.168.0.1
NIC1-2	10.0.0.1
NIC2-1	192.168.0.2
NIC2-2	10.0.0.2
RS-232C device	/dev/ttyS0

- Shared disk

Device name	/dev/sdb2
Mount point	/mnt/sdb2
File system	ext3

3.5.4 Hardware configuration of the mirror disk type cluster

The hardware configuration of the mirror disk in EXPRESSCLUSTER is described below.

Unlike the shared disk type, a network to copy the mirror disk data is necessary. In general, a network is used with NIC for internal communication in EXPRESSCLUSTER.

Mirror disks need to be separated from the operating system; however, they do not depend on a connection interface (IDE or SCSI.)

- Sample cluster environment with mirror disks used (When cluster partitions and data partitions are allocated to OS-installed disks)

  In the following configuration, free partitions of the OS-installed disks are used as cluster partitions and data partitions.

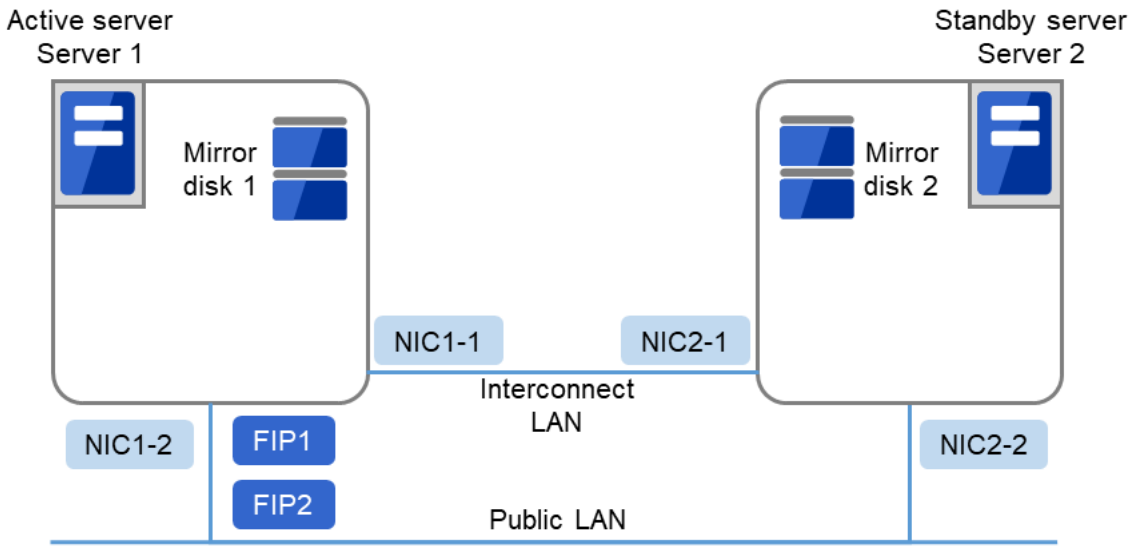


Fig. 3.17: Example of cluster configuration (1) (Mirror disk type)

FIP1	10.0.0.11 (Access destination from the Cluster WebUI client)
FIP2	10.0.0.12 (Access destination from the operation client)
NIC1-1	192.168.0.1
NIC1-2	10.0.0.1
NIC2-1	192.168.0.2
NIC2-2	10.0.0.2
RS-232C device	/dev/ttyS0

/boot device for OS	/dev/sda1
Swap device for OS	/dev/sda2
/(root) device for OS	/dev/sda3
Device for cluster partitions	/dev/sda5
Device for data partitions	/dev/sda6
Mount point	/mnt/sda6
File system	ext3

- Sample cluster environment with mirror disks used (When disks are prepared for cluster partitions and data partitions)

  In the following configuration, disks are prepared to be used for cluster partitions and data partitions, and connected to the servers.

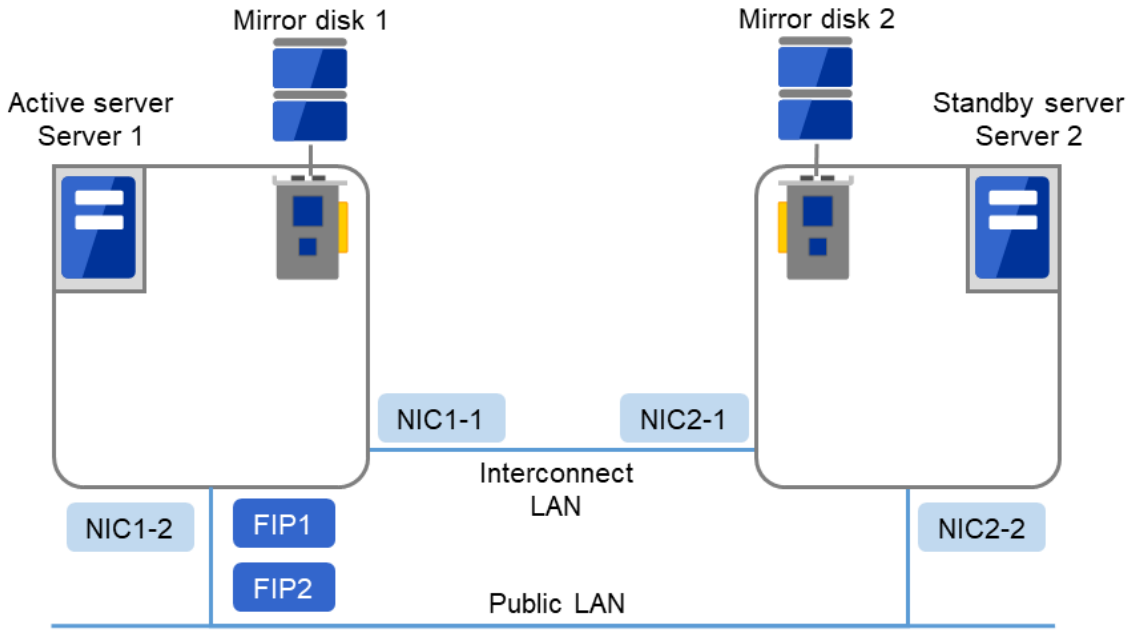


Fig. 3.18: Example of cluster configuration (2) (Mirror disk type)

FIP1	10.0.0.11 (Access destination from the Cluster WebUI client)
FIP2	10.0.0.12 (Access destination from the operation client)
NIC1-1	192.168.0.1
NIC1-2	10.0.0.1
NIC2-1	192.168.0.2
NIC2-2	10.0.0.2
RS-232C device	/dev/ttyS0

/boot device for OS	/dev/sda1
Swap device for OS	/dev/sda2
/(root) device for OS	/dev/sda3
Device for cluster partitions	/dev/sdb1
Mirror resource disk device	/dev/sdb2
Mount point	/mnt/sdb2
File system	ext3

3.5.5 Hardware configuration of the hybrid disk type cluster

The hardware configuration of the hybrid disk in EXPRESSCLUSTER is described below.

Unlike the shared disk type, a network to copy the data is necessary. In general, NIC for internal communication in EXPRESSCLUSTER is used to meet this purpose.

Disks do not depend on a connection interface (IDE or SCSI).

- Sample cluster environment with the hybrid disk used (When a shared disk is used by two servers and the data is mirrored to the normal disk of the third server)

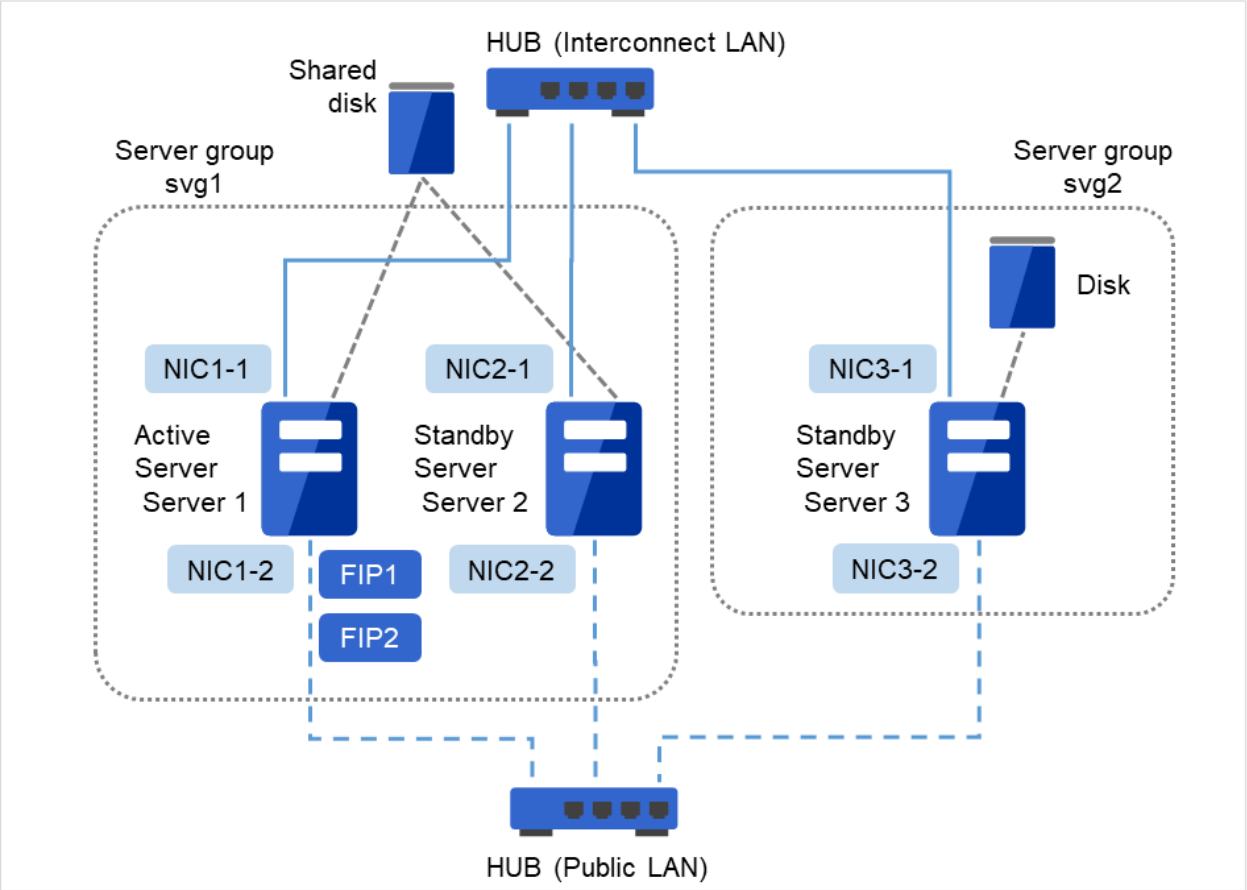


Fig. 3.19: Example of cluster configuration (Hybrid disk type)

FIP1	10.0.0.11 (Access destination from the Cluster WebUI client)
FIP2	10.0.0.12 (Access destination from the operation client)
NIC1-1	192.168.0.1
NIC1-2	10.0.0.1
NIC2-1	192.168.0.2
NIC2-2	10.0.0.2
NIC3-1	192.168.0.3
NIC3-2	10.0.0.3

- Shared disk

Hybrid device	/dev/NMP1
Mount point	/mnt/hd1
File system	ext3
Device for cluster partitions	/dev/sdb1
Hybrid resource disk device	/dev/sdb2
Disk heartbeat device name	/dev/sdb3
Raw device name	/dev/raw/raw1

- Disk for hybrid resource

Hybrid device	/dev/NMP1
Mount point	/mnt/hd1
File system	ext3
Device for cluster partitions	/dev/sdb1
Hybrid resource disk device	/dev/sdb2

3.5.6 What is cluster object?

In EXPRESSCLUSTER, the various resources are managed as the following groups:

- Cluster object
  Configuration unit of a cluster.
- Server object
  Indicates the physical server and belongs to the cluster object.
- Server group object
  Groups the servers and belongs to the cluster object.
- Heartbeat resource object
  Indicates the network part of the physical server and belongs to the server object.
- Network partition resolution resource object
  Indicates the network partition resolution mechanism and belongs to the server object.
- Group object
  Indicates a virtual server and belongs to the cluster object.
- Group resource object
  Indicates resources (network, disk) of the virtual server and belongs to the group object.
- Monitor resource object
  Indicates monitoring mechanism and belongs to the cluster object.

3.6 What is a resource?

In EXPRESSCLUSTER, a group used for monitoring the target is called "resources." There are four types of resources and are managed separately. Having resources allows distinguishing what is monitoring and what is being monitored more clearly. It also makes building a cluster and handling an error easy. The resources can be divided into heartbeat resources, network partition resolution resources, group resources, and monitor resources.

3.6.1 Heartbeat resources

Heartbeat resources are used for verifying whether the other server is working properly between servers. The following heartbeat resources are currently supported:

- LAN heartbeat resource
  Uses Ethernet for communication.
- Kernel mode LAN heartbeat resource
  Uses Ethernet for communication.
- Disk heartbeat resource
  Uses a specific partition (cluster partition for disk heartbeat) on the shared disk for communication. It can be used only on a shared disk configuration.
- Witness heartbeat resource
  Uses the external server running the Witness server service to show the status (of communication with each server) obtained from the external server.

3.6.2 Network partition resolution resources

The following resource is used to resolve a network partition.

- PING network partition resolution resource
  This is a network partition resolution resource by the PING method.
- HTTP network partition resolution resource
  This is a network partition resolution resource by the HTTP method.

3.6.3 Group resources

A group resource constitutes a unit when a failover occurs. The following group resources are currently supported:

- Floating IP resource (fip)
  Provides a virtual IP address. A client can access virtual IP address the same way as the regular IP address.
- EXEC resource (exec)
  Provides a mechanism for starting and stopping the applications such as DB and httpd.
- Disk resource (disk)
  Provides a specified partition on the shared disk. It can be used only on a shared disk configuration.
- Mirror disk resource (md)
  Provides a specified partition on the mirror disk. It can be used only on a mirror disk configuration.
- Hybrid disk resource (hd)
  Provides a specified partition on a shared disk or a disk. It can be used only for hybrid configuration.

- Volume manager resource (volmgr)
  Handles multiple storage devices and disks as a single logical disk.
- Virtual IP resource (vip)
  Provides a virtual IP address. This can be accessed from a client in the same way as a general IP address. This can be used in the remote cluster configuration among different network addresses.
- Dynamic DNS resource (ddns)
  Registers the virtual host name and the IP address of the active server to the dynamic DNS server.
- AWS elastic ip resource (awseip)
  Provides a system for giving an elastic IP (referred to as EIP) when EXPRESSCLUSTER is used on AWS.
- AWS virtual ip resource (awsvip)
  Provides a system for giving a virtual IP (referred to as VIP) when EXPRESSCLUSTER is used on AWS.
- AWS secondary ip resource (awssip)
  Provides a system for giving a secondary IP when EXPRESSCLUSTER is used on AWS.
- AWS DNS resource (awsdns)
  Registers the virtual host name and the IP address of the active server to Amazon Route 53 when EXPRESSCLUSTER is used on AWS.
- Azure probe port resource (azurepp)
  Provides a system for opening a specific port on a node on which the operation is performed when EXPRESSCLUSTER is used on Microsoft Azure.
- Azure DNS resource (azuredns)
  Registers the virtual host name and the IP address of the active server to Azure DNS when EXPRESSCLUSTER is used on Microsoft Azure.
- Google Cloud virtual IP resource (gcvip)
  Provides a system for opening a specific port on a node on which the operation is performed when EXPRESSCLUSTER is used on Google Cloud Platform.
- Google Cloud DNS resource (gcdns)
  Registers the virtual host name and the IP address of the active server to Cloud DNS when EXPRESSCLUSTER is used on Google Cloud Platform.
- Oracle Cloud virtual IP resource (ocvip)
  Provides a system for opening a specific port on a node on which the operation is performed when EXPRESSCLUSTER is used on Oracle Cloud Infrastructure.

3.6.4 Monitor resources

A monitor resource monitors a cluster system. The following monitor resources are currently supported:

- Floating IP monitor resource (fipw)
  Provides a monitoring mechanism of an IP address started up by a floating IP resource.
- IP monitor resource (ipw)
  Provides a monitoring mechanism of an external IP address.
- Disk monitor resource (diskw)
  Provides a monitoring mechanism of the disk. It also monitors the shared disk.
- Mirror disk monitor resource (mdw)
  Provides a monitoring mechanism of the mirroring disks.

- Mirror disk connect monitor resource (mdnw)
  Provides a monitoring mechanism of the mirror disk connect.
- Hybrid disk monitor resource (hdw)
  Provides a monitoring mechanism of the hybrid disk.
- Hybrid disk connect monitor resource (hdnw)
  Provides a monitoring mechanism of the hybrid disk connect.
- PID monitor resource (pidw)
  Provides a monitoring mechanism to check whether a process started up by exec resource is active or not.
- User mode monitor resource (userw)
  Provides a monitoring mechanism for a stalling problem in the user space.
- NIC Link Up/Down monitor resource (miiw)
  Provides a monitoring mechanism for link status of LAN cable.
- Volume manager monitor resource (volmgrw)
  Provides a monitoring mechanism for multiple storage devices and disks.
- Multi target monitor resource (mtw)
  Provides a status with multiple monitor resources.
- Virtual IP monitor resource (vipw)
  Provides a mechanism for sending RIP packets of a virtual IP resource.
- ARP monitor resource (arpw)
  Provides a mechanism for sending ARP packets of a floating IP resource or a virtual IP resource.
- Custom monitor resource (genw)
  Provides a monitoring mechanism to monitor the system by the operation result of commands or scripts which perform monitoring, if any.
- Message receive monitor resource (mrw)
  Specifies the action to take when an error message is received and how the message is displayed on the Cluster WebUI.
- Dynamic DNS monitor resource (ddnsw)
  Periodically registers the virtual host name and the IP address of the active server to the dynamic DNS server.
- Process name monitor resource (psw)
  Provides a monitoring mechanism for checking whether a process specified by a process name is active.
- DB2 monitor resource (db2w)
  Provides a monitoring mechanism for IBM DB2 database.
- FTP monitor resource (ftpw)
  Provides a monitoring mechanism for FTP server.
- HTTP monitor resource (httpw)
  Provides a monitoring mechanism for HTTP server.
- IMAP4 monitor resource (imap4w)
  Provides a monitoring mechanism for IMAP4 server.
- MySQL monitor resource (mysqlw)
  Provides a monitoring mechanism for MySQL database.
- NFS monitor resource (nfsw)

Provides a monitoring mechanism for nfs file server.

- Oracle monitor resource (oraclew)
  Provides a monitoring mechanism for Oracle database.
- Oracle Clusterware Synchronization Management monitor resource (osmw)
  Provides a monitoring mechanism for Oracle Clusterware process linked EXPRESSCLUSTER.
- POP3 monitor resource (pop3w)
  Provides a monitoring mechanism for POP3 server.
- PostgreSQL monitor resource (psqlw)
  Provides a monitoring mechanism for PostgreSQL database.
- Samba monitor resource (sambaw)
  Provides a monitoring mechanism for samba file server.
- SMTP monitor resource (smtpw)
  Provides a monitoring mechanism for SMTP server.
- Tuxedo monitor resource (tuxw)
  Provides a monitoring mechanism for Tuxedo application server.
- WebSphere monitor resource (wasw)
  Provides a monitoring mechanism for WebSphere application server.
- WebLogic monitor resource (wlsw)
  Provides a monitoring mechanism for WebLogic application server.
- WebOTX monitor resource (otxsw)
  Provides a monitoring mechanism for WebOTX application server.
- JVM monitor resource (jraw)
  Provides a monitoring mechanism for Java VM.
- System monitor resource (sraw)
  Provides a monitoring mechanism for the resources of the whole system.
- Process resource monitor resource(psrw)
  Provides a monitoring mechanism for running processes on the server.
- AWS Elastic IP monitor resource (awseipw)
  Provides a monitoring mechanism for the elastic ip given by the AWS elastic ip (referred to as EIP) resource.
- AWS Virtual IP monitor resource (awsvipw)
  Provides a monitoring mechanism for the virtual ip given by the AWS virtual ip (referred to as VIP) resource.
- AWS Secondary IP monitor resource (awssipw)
  Provides a monitoring mechanism for the secondary ip given by the AWS secondary ip resource.
- AWS AZ monitor resource (awsazw)
  Provides a monitoring mechanism for an Availability Zone (referred to as AZ).
- AWS DNS monitor resource (awsdnsw)
  Provides a monitoring mechanism for the virtual host name and IP address provided by the AWS DNS resource.
- Azure probe port monitor resource (azureppw)
  Provides a monitoring mechanism for probe port for the node where an Azure probe port resource has been activated.

- Azure load balance monitor resource (azurelbw)
  Provides a mechanism for monitoring whether the port number that is same as the probe port is open for the node where an Azure probe port resource has not been activated.
- Azure DNS monitor resource (azurednsw)
  Provides a monitoring mechanism for the virtual host name and IP address provided by the Azure DNS resource.
- Google Cloud virtual IP monitor resource (gcvipw)
  Provides a mechanism for monitoring the alive-monitoring port for the node where a Google Cloud virtual IP resource has been activated.
- Google Cloud load balance monitor resource (gclbw)
  Provides a mechanism for monitoring whether the same port number as the health-check port number has already been used , for the node where a Google Cloud virtual IP resource has not been activated.
- Google Cloud DNS monitor resource (gcdnsw)
  Provides a monitoring mechanism for the virtual host name and IP address provided by the Google Cloud DNS resource.
- Oracle Cloud virtual IP monitor resource (ocvipw)
  Provides a mechanism for monitoring the alive-monitoring port for the node where an Oracle Cloud virtual IP resource has been activated.
- Oracle Cloud load balance monitor resource (oclbw)
  Provides a mechanism for monitoring whether the same port number as the health-check port number has already been used , for the node where an Oracle Cloud virtual IP resource has not been activated.

3.7 Getting started with EXPRESSCLUSTER

Refer to the following guides when building a cluster system with EXPRESSCLUSTER:

3.7.1 Latest information

Refer to "4. *Installation requirements for EXPRESSCLUSTER*" and "5. *Latest version information*" and "6. *Notes and Restrictions*" and "7. *Upgrading EXPRESSCLUSTER*" in this guide.

3.7.2 Designing a cluster system

Refer to "Determining a system configuration" and "Configuring a cluster system" in the "Installation and Configuration Guide"; "Group resource details", "Monitor resource details", "Heartbeat resources details", "Network partition resolution resources details", and "Information on other settings" in the "Reference Guide" ; and the "Hardware Feature Guide".

3.7.3 Configuring a cluster system

Refer to the "Installation and Configuration Guide".

3.7.4 Troubleshooting the problem

Refer to "The system maintenance information" in the "Maintenance Guide", and "Troubleshooting" and "Error messages" in the "Reference Guide".

CHAPTER

FOUR

INSTALLATION REQUIREMENTS FOR EXPRESSCLUSTER

This chapter provides information on system requirements for EXPRESSCLUSTER.

This chapter covers:

4.1 Hardware

EXPRESSCLUSTER operates on the following server architectures:

- x86_64
- IBM POWER (Replicator, Replicator DR, Agents except Database Agent are not supported)
- IBM POWER LE (Replicator, Replicator DR and Agents are not supported)

4.1.1 General server requirements

Required specifications for EXPRESSCLUSTER Server are the following:

- RS-232C port 1 port (not necessary when configuring a cluster with 3 or more nodes)
- Ethernet port 2 or more ports
- Shared disk
- Mirror disk or empty partition for mirror
- DVD-ROM drive

4.2 Software

4.2.1 System requirements for EXPRESSCLUSTER Server

4.2.2 Supported distributions and kernel versions

The environment where EXPRESSCLUSTER Server can operate depends on kernel module versions because there are kernel modules unique to EXPRESSCLUSTER.

There are the following driver modules unique to EXPRESSCLUSTER.

Driver module unique to EXPRESSCLUSTER	Description
Kernel mode LAN heartbeat driver	Used with kernel mode LAN heartbeat resources.
Keepalive driver	Used if keepalive is selected as the monitoring method for user-mode monitor resources. Used if keepalive is selected as the monitoring method for shutdown monitoring.
Mirror driver	Used with mirror disk resources.

Kernel versions which has been verified are listed below.

About newest information, see the web site as follows:

EXPRESSCLUSTER website
->System Requirements
->EXPRESSCLUSTER X for Linux

Note: For the kernel version of Cent OS supported by EXPRESSCLUSTER, see the supported kernel version of Red Hat Enterprise Linux.

4.2.3 Applications supported by monitoring options

Version information of the applications to be monitored by monitor resources is described below.

x86_64

Monitor resource	Monitored application	EXPRESSCLUSTER version	Remarks
Oracle monitor	Oracle Database 19c (19.3)	5.0.0-1 or later	
DB2 monitor	DB2 V11.5	5.0.0-1 or later	
PostgreSQL monitor	PostgreSQL 14.1	5.0.0-1 or later	

Continued on next page

Table 4.2 – continued from previous page

Monitor resource	Monitored application	EXPRESSCLUSTER version	Remarks
	PowerGres on Linux 13.5	5.0.0-1 or later	
MySQL monitor	MySQL 8.0	5.0.0-1 or later	
	MariaDB 10.5	5.0.0-1 or later	
SQL Server monitor	SQL Server 2019	5.0.0-1 or later	
Samba monitor	Samba 3.3	4.0.0-1 or later	
	Samba 3.6	4.0.0-1 or later	
	Samba 4.0	4.0.0-1 or later	
	Samba 4.1	4.0.0-1 or later	
	Samba 4.2	4.0.0-1 or later	
	Samba 4.4	4.0.0-1 or later	
	Samba 4.6	4.0.0-1 or later	
	Samba 4.7	4.1.0-1 or later	
	Samba 4.8	4.1.0-1 or later	
	Samba 4.13	4.3.0-1 or later	
NFS monitor	nfsd 2 (udp)	4.0.0-1 or later	
	nfsd 3 (udp)	4.0.0-1 or later	
	nfsd 4 (tcp)	4.0.0-1 or later	
	mountd 1 (tcp)	4.0.0-1 or later	
	mountd 2 (tcp)	4.0.0-1 or later	
	mountd 3 (tcp)	4.0.0-1 or later	
HTTP monitor	No specified version	4.0.0-1 or later	
SMTP monitor	No specified version	4.0.0-1 or later	
POP3 monitor	No specified version	4.0.0-1 or later	
imap4 monitor	No specified version	4.0.0-1 or later	
ftp monitor	No specified version	4.0.0-1 or later	
Tuxedo monitor	Tuxedo 12c Release 2 (12.1.3)	4.0.0-1 or later	
WebLogic monitor	WebLogic Server 11g R1	4.0.0-1 or later	
	WebLogic Server 11g R2	4.0.0-1 or later	
	WebLogic Server 12c R2 (12.2.1)	4.0.0-1 or later	
	WebLogic Server 14c (14.1.1)	4.2.0-1 or later	
WebSphere monitor	WebSphere Application Server 8.5	4.0.0-1 or later	
	WebSphere Application Server 8.5.5	4.0.0-1 or later	
	WebSphere Application Server 9.0	4.0.0-1 or later	
WebOTX monitor	WebOTX Application Server V9.1	4.0.0-1 or later	
	WebOTX Application Server V9.2	4.0.0-1 or later	
	WebOTX Application Server V9.3	4.0.0-1 or later	
	WebOTX Application Server V9.4	4.0.0-1 or later	
	WebOTX Application Server V10.1	4.0.0-1 or later	
	WebOTX Application Server V10.3	4.3.0-1 or later	
JVM monitor	WebLogic Server 11g R1	4.0.0-1 or later	
	WebLogic Server 11g R2	4.0.0-1 or later	
	WebLogic Server 12c	4.0.0-1 or later	
	WebLogic Server 12c R2 (12.2.1)	4.0.0-1 or later	
	WebLogic Server 14c (14.1.1)	4.2.0-1 or later	
	WebOTX Application Server V9.1	4.0.0-1 or later	

Continued on next page

Table 4.2 – continued from previous page

Monitor resource	Monitored application	EXPRESSCLUSTER version	Remarks
	WebOTX Application Server V9.2	4.0.0-1 or later	WebOTX update is required to monitor process groups
	WebOTX Application Server V9.3	4.0.0-1 or later	
	WebOTX Application Server V9.4	4.0.0-1 or later	
	WebOTX Application Server V10.1	4.0.0-1 or later	
	WebOTX Application Server V10.3	4.3.0-1 or later	
	WebOTX Enterprise Service Bus V8.4	4.0.0-1 or later	
	WebOTX Enterprise Service Bus V8.5	4.0.0-1 or later	
	WebOTX Enterprise Service Bus V10.3	4.3.0-1 or later	
	JBoss Enterprise Application Platform 7.0	4.0.0-1 or later	
	JBoss Enterprise Application Platform 7.3	4.3.2-1 or later	
	JBoss Enterprise Application Platform 7.4	5.0.2-1 or later	
	Apache Tomcat 8.0	4.0.0-1 or later	
	Apache Tomcat 8.5	4.0.0-1 or later	
	Apache Tomcat 9.0	4.0.0-1 or later	
	Apache Tomcat 10.0	5.0.2-1 or later	
	WebSAM SVF for PDF 9.0	4.0.0-1 or later	
	WebSAM SVF for PDF 9.1	4.0.0-1 or later	
	WebSAM SVF for PDF 9.2	4.0.0-1 or later	
	WebSAM Report Director Enterprise 9.0	4.0.0-1 or later	
	WebSAM Report Director Enterprise 9.1	4.0.0-1 or later	
	WebSAM Report Director Enterprise 9.2	4.0.0-1 or later	
	WebSAM Universal Connect/X 9.0	4.0.0-1 or later	
	WebSAM Universal Connect/X 9.1	4.0.0-1 or later	
	WebSAM Universal Connect/X 9.2	4.0.0-1 or later	
System monitor	No specified version	4.0.0-1 or later	
Process resource monitor	No specified version	4.1.0-1 or later	

Note: To use monitoring options in x86_64 environments, applications to be monitored must be x86_64 version.

IBM POWER

Monitor resource	Monitored application	EXPRESSCLUSTER version	Remarks
DB2 monitor	DB2 V10.5	4.0.0-1 or later	
PostgreSQL monitor	PostgreSQL 9.3	4.0.0-1 or later	
	PostgreSQL 9.4	4.0.0-1 or later	
	PostgreSQL 9.5	4.0.0-1 or later	
	PostgreSQL 9.6	4.0.0-1 or later	
	PostgreSQL 10	4.0.0-1 or later	
	PostgreSQL 11	4.1.0-1 or later	

Note: To use monitoring options in IBM POWER environments, applications to be monitored must be IBM POWER version.

4.2.4 Operation environment for JVM monitor

The use of the JVM monitor requires a Java runtime environment. Also, monitoring a domain mode of JBoss Enterprise Application Platform requires Java(TM) SE Development Kit.

Java(TM) Runtime Environment	Version 7.0 Update 6 (1.7.0_6) or later
Java(TM) SE Development Kit	Version 7.0 Update 1 (1.7.0_1) or later
Java(TM) Runtime Environment	Version 8.0 Update 11 (1.8.0_11) or later
Java(TM) SE Development Kit	Version 8.0 Update 11 (1.8.0_11) or later
Java(TM) Runtime Environment	Version 9.0 (9.0.1) or later
Java(TM) SE Development Kit	Version 9.0 (9.0.1) or later
Java(TM) SE Development Kit	Version 11.0 (11.0.5) or later
Open JDK	Version 7.0 Update 45 (1.7.0_45) or later Version 8.0 (1.8.0) or later Version 9.0 (9.0.1) or later

4.2.5 Operation environment for AWS elastic ip resource, AWS virtual ip resource, AWS Elastic IP monitor resource, AWS virtual IP monitor resource, AWS AZ monitor resource

The use of the AWS elastic ip resource, AWS virtual ip resource, AWS Elastic IP monitor resource, AWS virtual IP monitor resource, AWS AZ monitor resource requires the following software.

Software	Version	Remarks
AWS CLI	1.6.0 or later 2.0.0 or later	

Continued on next page

Table 4.5 – continued from previous page

Software	Version	Remarks
Python	2.6.5 or later 2.7.5 or later 3.5.2 or later 3.6.8 or later 3.8.1 or later 3.8.3 or later	Python accompanying the AWS CLI is not allowed.

The environment where EXPRESSCLUSTER Server can operate depends on kernel module versions because there are kernel modules unique to EXPRESSCLUSTER.
Since the OS is frequently upgraded that AWS is to provide, when it is not possible to behavior will occur.
Kernel versions which has been verified, please refer to the "4.2.2. *Supported distributions and kernel versions*".

4.2.6 Operation environment for AWS secondary ip resource, AWS secondary IP monitor resource

The use of the AWS secondary ip resource, AWS secondary IP monitor resource requires the following software.

Software	Version	Remarks
AWS CLI	2.0.0 or later	

The environment where EXPRESSCLUSTER Server can operate depends on kernel module versions because there are kernel modules unique to EXPRESSCLUSTER.
Since the OS is frequently upgraded that AWS is to provide, when it is not possible to behavior will occur.
Kernel versions which has been verified, please refer to the "4.2.2. *Supported distributions and kernel versions*".

4.2.7 Operation environment for AWS DNS resource, AWS DNS monitor resource

The use of the AWS DNS resource, AWS DNS monitor resource requires the following software.

Software	Version	Remarks
AWS CLI	1.11.0 or later	
Python (When OS is Red Hat Enterprise Linux 6, Cent OS 6, SUSE Linux Enterprise Server 11, Oracle Linux 6)	2.6.6 or later 3.6.5 or later 3.8.1 or later	Python accompanying the AWS CLI is not allowed.

Continued on next page

Table 4.7 – continued from previous page

Software	Version	Remarks
Python (When OS is besides Red Hat Enterprise Linux 6, Cent OS 6, SUSE Linux Enterprise Server 11, Oracle Linux 6)	2.7.5 or later 3.5.2 or later 3.6.8 or later 3.8.1 or later 3.8.3 or later	Python accompanying the AWS CLI is not allowed.

The environment where EXPRESSCLUSTER Server can operate depends on kernel module versions because there are kernel modules unique to EXPRESSCLUSTER.
Since the OS is frequently upgraded that AWS is to provide, when it is not possible to behavior will occur.
Kernel versions which has been verified, please refer to the "4.2.2. *Supported distributions and kernel versions*".

4.2.8 Operation environment for AWS forced stop resource

The use of the AWS forced stop resource requires the following software.

Software	Version	Remarks
AWS CLI	2.0.0 or later	

The environment where EXPRESSCLUSTER Server can operate depends on kernel module versions because there are kernel modules unique to EXPRESSCLUSTER.
Since the OS is frequently upgraded that AWS is to provide, when it is not possible to behavior will occur.
Kernel versions which has been verified, please refer to the "4.2.2. *Supported distributions and kernel versions*".

4.2.9 Operation environment for Azure probe port resource, Azure probe port monitor resource, Azure load balance monitor resource

The environment where EXPRESSCLUSTER Server can operate depends on kernel module versions because there are kernel modules unique to EXPRESSCLUSTER.
Since the OS is frequently upgraded that Microsoft Azure is to provide, when it is not possible to behavior will occur.
Kernel versions which has been verified, please refer to the "4.2.2. *Supported distributions and kernel versions*".

The following are the Microsoft Azure deployment models with which the operation of the Azure probe port resource is verified. For details on how to set up a Load Balancer, refer to the documents from Microsoft (https://azure.microsoft.com/en-us/documentation/articles/load-balancer-arm/).

x86_64

Deployment model	EXPRESSCLUSTER Version	Remark
Resource Manager	4.0.0-1 or later	Load balancer is required

4.2.10 Operation environment for Azure DNS resource, Azure DNS monitor resource

The use of the Azure DNS resource, Azure DNS monitor resource requires the following software.

Software	Version	Remarks
Azure CLI (When OS is Red Hat Enterprise Linux 6, Cent OS 6, Asianux Server 4, SUSE Linux Enterprise Server 11, Oracle Linux 6)	1.0 or later	Python is not required.
Azure CLI (When OS is besides Red Hat Enterprise Linux 6, Cent OS 6, Asianux Server 4, SUSE Linux Enterprise Server 11, Oracle Linux 6)	2.0 or later	

Using Azure CLI 2.0 is recommended because Azure CLI 1.0 (the Azure classic CLI) is no longer recommended. For more information, refer to the following:

Differences between Azure CLI products:
https://docs.microsoft.com/en-us/cli/azure/install-classic-cli?view=azure-cli-latest

For the prerequisites of the Azure CLI and the instructions on how to install it, refer to the following:

Install the Azure CLI:
https://docs.microsoft.com/en-us/cli/azure/install-azure-cli?view=azure-cli-latest

The environment where EXPRESSCLUSTER Server can operate depends on kernel module versions because there are kernel modules unique to EXPRESSCLUSTER.
Since the OS is frequently upgraded that Microsoft Azure is to provide, when it is not possible to behavior will occur.
Kernel versions which has been verified, please refer to the "4.2.2. *Supported distributions and kernel versions*".

The following are the Microsoft Azure deployment models with which the operation of the Azure DNS resource, the Azure DNS monitor resource is verified. For setting about Azure DNS, please refer to the "EXPRESSCLUSTER X HA Cluster Configuration Guide for Microsoft Azure (Linux)"

x86_64

Deployment model	EXPRESSCLUSTER Version	Remark
Resource Manager	4.0.0-1 or later	Azure DNS is required

4.2.11 Operation environments for Google Cloud virtual IP resource, Google Cloud virtual IP monitor resource, and Google Cloud load balance monitor resource

The operation environment for EXPRESSCLUSTER Server depends on the kernel module version because there are kernel modules unique to EXPRESSCLUSTER.

Frequently upgraded, the OS on Google Cloud Platform may fail to operate.

For information on verified kernel versions, refer to "4.2.2. *Supported distributions and kernel versions*".

4.2.12 Operation environments for Google Cloud DNS resource, Google Cloud DNS monitor resource

The use of the Google Cloud DNS resource, Azure Google Cloud monitor resource requires the following software.

Software	Version	Remarks
Google Cloud SDK	295.0.0~	

For the prerequisites of the Google Cloud SDK and the instructions on how to install it, refer to the following:

Install the Google Cloud SDK:
https://cloud.google.com/sdk/install

The environment where EXPRESSCLUSTER Server can operate depends on kernel module versions because there are kernel modules unique to EXPRESSCLUSTER.

Since the OS is frequently upgraded that Google Cloud Platform is to provide, when it is not possible to behavior will occur.

Kernel versions which has been verified, please refer to the "4.2.2. *Supported distributions and kernel versions*".

4.2.13 Operation environments for Oracle Cloud virtual IP resource, Oracle Cloud virtual IP monitor resource, and Oracle Cloud load balance monitor resource

The operation environment for EXPRESSCLUSTER Server depends on the kernel module version because there are kernel modules unique to EXPRESSCLUSTER.

Frequently upgraded, the OS on Oracle Cloud Infrastructure may fail to operate.

For information on verified kernel versions, refer to "4.2.2. *Supported distributions and kernel versions*".

4.2.14 Operation environment for OCI forced stop resource

The use of the OCI forced stop resource requires the following software.

Software	Version	Remarks
OCI CLI	3.5.3 or later	

The environment where EXPRESSCLUSTER Server can operate depends on kernel module versions because there are kernel modules unique to EXPRESSCLUSTER.
Since the OS is frequently upgraded that OCI is to provide, when it is not possible to behavior will occur.
Kernel versions which has been verified, please refer to the "4.2.2. *Supported distributions and kernel versions*".

4.2.15 Operation environment for clpcfadm.py command

Using the clpcfadm.py command requires the following software:

Software	Version	Remarks
Python	3.6.8~	

4.2.16 Required memory and disk size

Required memory size (User mode)	200MB[2]
Required memory size (Kernel mode)	When the synchronization mode is used: 1MB + (number of request queues x I/O size) + (2MB + Difference Bitmap Size x number of mirror disk resources and hybrid disk resources When the asynchronous mode is used: 1MB + (number of request queues x I/O size) + (3MB + (number of asynchronous queues x I/O size) + (I/O size / 4KB x 8B + 0.5KB) x (max size of history file / I/O size + number of asynchronous queues) + (Difference Bitmap Size)) x number of mirror disk resources and hybrid disk resources When the kernel mode LAN heartbeat driver is used: 8MB When the keepalive driver is used: 8MB
Required disk size (Right after installation)	300MB

Continued on next page

Table 4.15 – continued from previous page

Required disk size (During operation)	5.0GB + 1.0GB[3]

Note: Estimated I/O size is as follows:

- 2MB (RHEL8)
- 1MB (Ubuntu16)
- 124KB (Ubuntu14, RHEL7)
- 4KB (RHEL6)

For the setting value of the number of request queues and asynchronization queues, see "Understanding Mirror disk resources" of "Group resource details" in the "Reference Guide".

For the required size of a partition for a disk heartbeat resource, see "*Shared disk*".

For the required size of a cluster partition, see "*Mirror disk*" and "*Hybrid disk*".

[2] excepting for optional products.
[3] A disk capacity required to use mirror disk resources and hybrid disk resources.

4.3 System requirements for the Cluster WebUI

4.3.1 Supported operating systems and browsers

Refer to the website, *http://www.nec.com/global/prod/expresscluster/,* for the latest information. Currently the following operating systems and browsers are supported:

Browser	Language
Internet Explorer 11	English/Japanese/Chinese
Internet Explorer 10	English/Japanese/Chinese
Firefox	English/Japanese/Chinese
Google Chrome	English/Japanese/Chinese
Microsoft Edge (Chromium)	English/Japanese/Chinese

Note: When using an IP address to connect to Cluster WebUI, the IP address must be registered to **Site** of **Local Intranet** in advance.

Note: When accessing Cluster WebUI with Internet Explorer 11, the Internet Explorer may stop with an error. In order to avoid it, please upgrade the Internet Explorer into KB4052978 or later. Additionally, in order to apply KB4052978 or later to Windows 8.1/Windows Server 2012R2, apply KB2919355 in advance. For details, see the information released by Microsoft.

Note: No mobile devices, such as tablets and smartphones, are supported.

4.3.2 Required memory and disk size

- Required memory size: 500 MB or more
- Required disk size: 200 MB or more

CHAPTER

FIVE

LATEST VERSION INFORMATION

This chapter provides the latest information on EXPRESSCLUSTER.

This chapter covers:

5.1 Correspondence list of EXPRESSCLUSTER and a manual

Description in this manual assumes the following version of EXPRESSCLUSTER. Make sure to note and check how EXPRESSCLUSTER versions and the editions of the manuals are corresponding.

EXPRESSCLUSTER Internal Version	Manual	Edition	Remarks
5.0.2-1	Getting Started Guide	4th Edition	
	Installation and Configuration Guide	2nd Edition	
	Reference Guide	4th Edition	
	Maintenance Guide	2nd Edition	
	Hardware Feature Guide	1st Edition	

5.2 New features and improvements

The following features and improvements have been released.

No.	Internal Version	Contents
1	5.0.0-1	The newly released kernel is now supported.
2	5.0.0-1	Ubuntu 20.04.3 LTS is now supported.
3	5.0.0-1	SUSE LINUX Enterprise Server 12 SP3 is now supported.
4	5.0.0-1	Along with the major upgrade, some functions have been removed. For details, refer to the list of removed functions.
5	5.0.0-1	Added a function to suppress the automatic failover against a server crash, collectively in the whole cluster.
6	5.0.0-1	Added a function to give a notice in an alert log that the server restart count was reset as the final action against the detected activation error or deactivation error of a group resource or against the detected error of a monitor resource.
7	5.0.0-1	Added a function to exclude a server (with an error detected by a specified monitor resource) from the failover destination, for the automatic failover other than dynamic failover.
8	5.0.0-1	Added the clpfwctrl command for adding a firewall rule.
9	5.0.0-1	Added AWS secondary IP resources and AWS secondary IP monitor resources.
10	5.0.0-1	The forced stop function using BMC has been redesigned as a BMC forced-stop resource.
11	5.0.0-1	Redesigned the function for forcibly stopping virtual machines as a vCenter forced-stop resource.
12	5.0.0-1	The forced stop function in the AWS environment has been added to forced stop resources.
13	5.0.0-1	The forced stop function in the OCI environment has been added to forced stop resources.
14	5.0.0-1	Redesigned the forced stop script as a custom forced-stop resource.
15	5.0.0-1	Added a function to collectively change actions (followed by OS shutdowns such as a recovery action following an error detected by a monitor resource) into OS reboots.
16	5.0.0-1	Improved the alert message regarding the wait process for start/stop between groups.
17	5.0.0-1	The display option for the clpstat configuration information has allowed displaying the setting value of the resource start attribute.
18	5.0.0-1	The clpcl/clpstdn command has allowed specifying the -h option even when the local server belongs to a stopped cluster.
19	5.0.0-1	A warning message is now displayed when Cluster WebUI is connected via a non-actual IP address and is switched to config mode.
20	5.0.0-1	In the config mode of Cluster WebUI, a group can now be deleted with the group resource registered.
21	5.0.0-1	Changed the content of the error message that a communication timeout occurred in Cluster WebUI.
22	5.0.0-1	Changed the content of the error message that executing the full copy failed on the mirror disk screen in Cluster WebUI.
23	5.0.0-1	Added a function to copy a group, group resource, or monitor resource registered in the config mode of Cluster WebUI.

Continued on next page

Table 5.2 – continued from previous page

No.	Internal Version	Contents
24	5.0.0-1	Added a function to move a group resource registered in the config mode of Cluster WebUI, to another group.
25	5.0.0-1	The settings can now be changed at the group resource list of [Group Properties] in the config mode of Cluster WebUI.
26	5.0.0-1	The settings can now be changed at the monitor resource list of [Monitor Common Properties] in the config mode of Cluster WebUI.
27	5.0.0-1	The dependency during group resource deactivation is now displayed in the config mode of Cluster WebUI.
28	5.0.0-1	Added a function to display a dependency diagram at the time of group resource activation/deactivation in the config mode of Cluster WebUI.
29	5.0.0-1	Added a function to narrow down a range of display by type or resource name of a group resource or monitor resource on the status screen of Cluster WebUI.
30	5.0.0-1	User mode monitor resources and dynamic DNS monitor resources now support the function for collecting cluster statistics information.
31	5.0.0-1	An intermediate certificate can now be used as a certificate file when HTTPS is used for communication in the WebManager service.
32	5.0.0-1	Added the clpcfconv.sh command, which changes the cluster configuration data file from the old version to the current one.
33	5.0.0-1	Added a function to delay the start of the cluster service for starting the OS.
34	5.0.0-1	Increased the items of cluster configuration data to be checked.
35	5.0.0-1	Details such as measures can now be displayed for error results of checking cluster configuration data in Cluster WebUI.
36	5.0.0-1	The OS type can be specified for specifying the create option of the clpcfset command.
37	5.0.0-1	Added a function to delete a resource or parameter from cluster configuration data, which is enabled by adding the del option to the clpcfset command.
38	5.0.0-1	Added the clpcfadm.py command, which enhances the interface for the clpcfset command.
39	5.0.0-1	The start completion timing of an AWS DNS resource has been changed to the timing before which the following is confirmed: The record set was propagated to AWS Route 53.
40	5.0.0-1	Changed the default value for [Wait Time to Start Monitoring] of AWS DNS monitor resources to 300 seconds.
41	5.0.0-1	Improved the functionality of monitor resources not to be affected by disk IO delay as follows: When a timeout occurs due to the disk wait dormancy (D state) of the monitor process, they consider the status as a warning instead of an error.
42	5.0.0-1	The clpstat command can now be run duplicately.
43	5.0.0-1	Added the Node Manager service.
44	5.0.0-1	Added a function for statistical information on heartbeat.
45	5.0.0-1	The proxy server has become available even when a Witness heartbeat resource is not used for an HTTP NP resolution resource.
46	5.0.0-1	SELinux enforcing mode is now supported.
47	5.0.0-1	HTTP monitor resources now support digest authentication.
48	5.0.0-1	The FTP server that uses FTPS for the FTP monitor resource can now be monitored.

Continued on next page

Table 5.2 – continued from previous page

No.	Internal Version	Contents
49	5.0.0-1	JBoss EAP domain mode of JVM monitor resources can now be monitored in Java 9 or later.
50	5.0.2-1	JVM monitor resource now supports JBoss Enterprise Application Platform 7.4.
51	5.0.2-1	JVM monitor resource supports Apache Tomcat 10.0.

5.3 Corrected information

Modification has been performed on the following minor versions.

Critical level:

L

Operation may stop. Data destruction or mirror inconsistency may occur.
Setup may not be executable.

M

Operation stop should be planned for recovery.
The system may stop if duplicated with another fault.

S

A matter of displaying messages.
Recovery can be made without stopping the system.

No.	Version in which the problem has been solved / Version in which the problem occurred	Phenomenon	Level	Occurrence condition/ Occurrence frequency
1	5.0.0-1 / 1.0.0-1 to 4.3.2-1	In a group, when a group resource alone is successfully activated, the restoration of another group resource may be executed.	S	This problem occurs in a group where a group resource alone is activated with another group resource failing in activation.
2	5.0.0-1 / 4.1.0-1 to 4.3.2-1	In the config mode of Cluster WebUI, modifying a comment on a group resource may not be applied.	S	This problem occurs in the following case: A comment on a group resource is modified, the [Apply] button is clicked, the change is undone, and then the [OK] button is clicked.
3	5.0.0-1 / 4.1.0-1 to 4.3.2-1	In the config mode of Cluster WebUI, modifying a comment on a monitor resource may not be applied.	S	This problem occurs in the following case: A comment on a monitor resource is modified, the [Apply] button is clicked, the change is undone, and then the [OK] button is clicked.
4	5.0.0-1 / 4.0.0-1 to 4.3.2-1	In the status screen of Cluster WebUI, a communication timeout during the operation of a cluster causes a request to be repeatedly issued.	M	This problem always occurs when a communication timeout occurs between Cluster WebUI and a cluster server.

Continued on next page

Table 5.3 – continued from previous page

No.	Version in which the problem has been solved / Version in which the problem occurred	Phenomenon	Level	Occurrence condition/ Occurrence frequency
5	5.0.0-1 / 4.1.0-1 to 4.3.2-1	Custer WebUI may freeze when dependency is set in the config mode of Cluster WebUI.	S	This problem occurs when two group resources are made dependent on each other.
6	5.0.0-1 / 4.2.0-1 to 4.3.2-1	The response of the clpstat command may be delayed.	S	This problem may occur when communication with other servers is cut off.
7	5.0.0-1 / 3.1.0-1 to 4.3.2-1	A cluster service may not stop.	S	This problem very rarely occurs when stopping a cluster service is tried.
8	5.0.0-1 / 4.0.0-1 to 4.3.2-1	A monitor resource may mistakenly detect a monitoring timeout.	M	This problem very rarely occurs when a monitoring process is executed by a monitor resource.
9	5.0.0-1 / 4.2.0-1 to 4.3.2-1	Executing the clpcfchk command as follows causes a mixture of the current check results and the previous ones: The -o option is used to specify a directory where a file of the previous check results exists.	S	This problem occurs when a directory specified with the -o option of the clpcfchk command includes a file of the previous check results (cfchk_result.csv).
10	5.0.0-1 / 4.3.0-1 to 4.3.2-1	In checking a cluster configuration, a check for fstab may fail.	S	This problem occurs with a slash (/) placed after a device name or mount point written into the /etc/fstab file.
11	5.0.0-1 / 4.3.0-1 to 4.3.2-1	The clpcfset command may abend.	S	This problem occurs when an empty string is specified as an attribute value.

Continued on next page

Table 5.3 – continued from previous page

No.	Version in which the problem has been solved / Version in which the problem occurred	Phenomenon	Level	Occurrence condition/ Occurrence frequency
12	5.0.0-1 / 4.0.0-1 to 4.3.2-1	In an AWS environment, a forced stop script may time out.	S	This problem may occur when a forced stop script is run In an AWS environment.
13	5.0.0-1 / 4.2.0-1 to 4.3.2-1	An error occurs when the status code of a target response is 301 in an HTTP NP resolution resource.	S	This problem occurs when the response status code is 301.
14	5.0.0-1 / 4.0.0-1 to 4.3.2-1	In the WebManager service, [Client Session Timeout] may not work.	S	This problem occurs in the following case: Before the time specified in [Client Session Timeout] passes, the next request is not issued.
15	5.0.0-1 / 4.0.0-1 to 4.3.2-1	When a hybrid disk resource is used, there is a difference in the order of listing servers between the status screen of and the mirror disk screen of Cluster WebUI.	S	This problem occurs depending on the server group name: While the status screen lists servers in the order of priority, the mirror disk screen lists them in the ascending order of the names of server groups to which the servers belong.
16	5.0.0-1 / 4.0.0-1 to 4.3.2-1	When a monitoring process by a monitor resource times out, detecting a monitoring error may take time.	S	This problem very rarely occurs when a monitoring process by a monitor resource times out.
17	5.0.0-1 / 1.0.0-1 to 4.3.2-1	With an IP monitor resource or PING NP resolution resource configured, a large number of ICMP packets may be sent.	S	This problem occurs when an ICMP communication process leads to receiving an unexpected packet.
18	5.0.0-1 / 4.0.0-1 to 4.3.2-1	In [Monitoring usage of memory] for process resource monitor resources, [Duration time (min)] has been replaced with [Maximum Refresh Count (time)].	S	This problem occurs when the properties are displayed with Cluster WebUI or the clpstat command.

Continued on next page

Table 5.3 – continued from previous page

No.	Version in which the problem has been solved / Version in which the problem occurred	Phenomenon	Level	Occurrence condition/ Occurrence frequency
19	5.0.0-1 / 3.3.2-1 to 4.3.2-1	With a mirror disk connection disconnected, a response to a command for mirror disks may be delayed.	S	This problem occurs when the most prioritized mirror disk connection of all is disconnected.
20	5.0.0-1 / 1.0.0-1 to 4.3.2-1	Deactivating a disk resource may fail with its disk type set to [raw].	S	This problem occurs in the following case: During the deactivation of a disk resource with its disk type set to [raw], a process exists accessing the device.
21	5.0.0-1 / 1.1.0-1 to 4.3.2-1	With a mirror disk connection disconnected, the OS may intermittently stall.	S	This problem may occur in the following case: A mirror disk resource or hybrid disk resource is in [Asynchronous] mode, and the I/O load is high with a mirror disk connection disconnected.
22	5.0.0-1 / 4.2.0-1 to 4.3.2-1	The EXPRESSCLUSTER Information Base service may abend.	S	This problem very rarely occurs when one of the following is performed: - Cluster startup - Cluster stop - Cluster suspension - Cluster resumption
23	5.0.1-1 / 5.0.0-1	In Ubuntu environments, the clpcfconv.sh command (for converting cluster configuration data files) fails to be executed.	S	This problem occurs in Ubuntu environments.
24	5.0.1-1 / 5.0.0-1	For a cluster configuration data file created on EXPRESSCLUSTER X 3.3 for Linux: After the data file is converted with the conversion command and then applied to the cluster, the mirror agent fails to start up.	S	This problem occurs when a mirror disk resource or hybrid disk resource is used with EXPRESSCLUSTER upgraded from X 3.3 for Linux.
25	5.0.1-1 / 5.0.0-1	For the clprexec command, the --script option does not work.	S	This problem occurs when the clprexec command is executed with the --script option specified.

Continued on next page

Table 5.3 – continued from previous page

No.	Version in which the problem has been solved / Version in which the problem occurred	Phenomenon	Level	Occurrence condition/ Occurrence frequency
26	5.0.1-1 / 5.0.0-1	After a forced-stop resource is added by executing the clpcfset command, the cluster fails to start up.	S	This problem occurs during an attempt to start up a cluster to which cluster configuration data (including a forced-stop resource added by executing the clpcfset command) was applied.
27	5.0.1-1 / 5.0.0-1	In Amazon Linux 2 environments, kernel mode LAN heartbeats do not start up normally.	M	This problem occurs in Amazon Linux 2 environments.
28	5.0.1-1 / 4.3.0-1 to 4.3.2-1 , 5.0.0-1	For mirror disk resources/hybrid disk resources based on the ext4 file system: A mirror recovery in full-copy mode may not normally copy data to the destination.	L	This problem occurs during a mirror recovery in full-copy mode with a mirror disk resource/hybrid disk resource based on the ext4 file system.
29	5.0.1-1 / 4.3.2-1 , 5.0.0-1	For Oracle monitor resources: When the monitoring times out, the retrying process may not work normally.	M	This problem occurs with an Oracle monitor resource when the monitoring process times out.
30	5.0.2-1 / 5.0.0-1 to 5.0.1-1	The Amazon CloudWatch linkage function may not work.	S	This problem occurs on very rare occasions with the Amazon CloudWatch linkage function configured.
31	5.0.2-1 / 5.0.0-1 to 5.0.1-1	A monitor resource may detect a monitoring timeout by mistake.	S	This problem occurs on very rare occasions during a monitoring process by the monitor resource.
32	5.0.2-1 / 1.0.0-1 to 5.0.1-1	Performing the keepalive reset and keepalive panic may fail.	S	This problem occurs when the major number (10) and the minor number (241), both of which should be used by the keepalive driver, are used by another driver.
33	5.0.2-1 / 4.3.0-1 to 5.0.1-1	The monitoring process of a Tuxedo monitor resource may abend, leading to a monitoring error.	M	The occurrence of this problem depends on the timing.

Continued on next page

Table 5.3 – continued from previous page

No.	Version in which the problem has been solved / Version in which the problem occurred	Phenomenon	Level	Occurrence condition/ Occurrence frequency
34	5.0.2-1 / 5.0.0-1 to 5.0.1-1	Forcibly stopping more than one server may fail.	S	This problem occurs on rare occasions when one of three or more servers in a cluster tries to forcibly stop other servers.
35	5.0.2-1 / 1.0.0-1 to 5.0.1-1	The clpstat command may abend.	S	This problem occurs in an environment where a failover group is set with no group resources registered.
36	5.0.2-1 / 5.0.0-1 to 5.0.1-1	With a cluster suspended, Cluster WebUI or the clpstat command may show the server status as stopped.	S	This problem occurs when both of the following services are restarted with the cluster suspended: - clusterpro_nm - clusterpro_ib
37	5.0.2-1 / 5.0.0-1 to 5.0.1-1	A group/monitor resource status may be incorrectly shown.	S	This problem occurs with something wrong in the internal processing of cluster services during OS startup.
38	5.0.2-1 / 5.0.0-1 to 5.0.1-1	Cluster WebUI or the clpstat command incorrectly shows the status of a server using no forced-stop resources.	S	This problem occurs when any of three or more servers in a cluster is configured not to use the forced-stop function.
39	5.0.2-1 / 5.0.0-1 to 5.0.1-1	Cluster WebUI displays the setting items of a high-speed SSD which does not work with an OS, one of the system requirements for EXPRESSCLUSTER X 5.0.	S	The setting items are always displayed in the detailed properties of mirror disk resources and in those of hybrid disk resources.
40	5.0.2-1 / 4.3.0-1 to 5.0.1-1	The clpwebmc process may abend.	S	This problem occurs on very rare occasions during cluster operation.

Continued on next page

Table 5.3 – continued from previous page

No.	Version in which the problem has been solved / Version in which the problem occurred	Phenomenon	Level	Occurrence condition/ Occurrence frequency
41	5.0.2-1 / 4.3.0-1 to 5.0.1-1	If the mount point of a disk resource, mirror disk resource, or hybrid disk resource includes a space, the /etc/fstab entry check (a function of checking cluster configuration data) fails.	S	This problem occurs when cluster configuration data is checked with the mount point including a space.

CHAPTER

SIX

NOTES AND RESTRICTIONS

This chapter provides information on known problems and how to troubleshoot the problems.

This chapter covers:

6.1 Designing a system configuration

Hardware selection, option products license arrangement, system configuration, and shared disk configuration are introduced in this section.

6.1.1 Function list and necessary license

The following option products are necessary as many as the number of servers.

> Those resources and monitor resources for which the necessary licenses are not registered are not on the resource list of the Cluster WebUI.

Necessary function	Necessary license
Mirror disk resource	EXPRESSCLUSTER X Replicator 5.0[4]
Hybrid disk resource	EXPRESSCLUSTER X Replicator DR 5.0[5]
Oracle monitor resource	EXPRESSCLUSTER X Database Agent 5.0
DB2 monitor resource	EXPRESSCLUSTER X Database Agent 5.0
PostgreSQL monitor resource	EXPRESSCLUSTER X Database Agent 5.0
MySQL monitor resource	EXPRESSCLUSTER X Database Agent 5.0
SQL Server monitor resource	EXPRESSCLUSTER X Database Agent 5.0
ODBC monitor resource	EXPRESSCLUSTER X Database Agent 5.0
Samba monitor resource	EXPRESSCLUSTER X File Server Agent 5.0
nfs monitor resource	EXPRESSCLUSTER X File Server Agent 5.0
http monitor resource	EXPRESSCLUSTER X Internet Server Agent 5.0
smtp monitor resource	EXPRESSCLUSTER X Internet Server Agent 5.0
pop3 monitor resource	EXPRESSCLUSTER X Internet Server Agent 5.0
imap4 monitor resource	EXPRESSCLUSTER X Internet Server Agent 5.0
ftp monitor resource	EXPRESSCLUSTER X Internet Server Agent 5.0
Tuxedo monitor resource	EXPRESSCLUSTER X Application Server Agent 5.0
WebLogic monitor resource	EXPRESSCLUSTER X Application Server Agent 5.0
WebSphere monitor resource	EXPRESSCLUSTER X Application Server Agent 5.0
WebOTX monitor resource	EXPRESSCLUSTER X Application Server Agent 5.0
JVM monitor resource	EXPRESSCLUSTER X Java Resource Agent 5.0
System monitor resource	EXPRESSCLUSTER X System Resource Agent 5.0
Process resource monitor resource	EXPRESSCLUSTER X System Resource Agent 5.0
Mail report actions	EXPRESSCLUSTER X Alert Service 5.0
Network Warning Light status	EXPRESSCLUSTER X Alert Service 5.0

6.1.2 Hardware requirements for mirror disks

- Linux md stripe set, volume set, mirroring, and stripe set with parity cannot be used for either mirror disk resource cluster partitions or data partitions.
- Linux LVM volumes can be used for both cluster partitions and data partitions.
  For SuSE, however, LVM and MultiPath volumes cannot be used for data partitions. (This is because for SuSE, ReadOnly or ReadWrite control over these volumes cannot be performed by EXPRESSCLUSTER.)
- Mirror disk resource cannot be made as a target of a Linux md stripe set, volume set, mirroring, and stripe set with parity.

[4] When configuring data mirror form, product **Replicator** must be purchased.
[5] When configuring hybrid disk form, product **Replicator DR** must be purchased.

- Mirror partitions (data partition and cluster partition) to use a mirror disk resource.
- There are two ways to allocate mirror partitions:
  - Allocate a mirror partition (data partition and cluster partition) on the disk where the operating system (such as root partition and swap partition) resides.
  - Reserve (or add) a disk (or LUN) not used by the operating system and allocate a mirror partition on the disk.
- Consider the following when allocating mirror partitions:
  - When maintainability and performance are important:
    - It is recommended to have a mirror disk that is not used by the OS.
  - When LUN cannot be added due to hardware RAID specification or when changing LUN configuration is difficult in hardware RAID pre-install model:
    - Allocate a mirror partition on the same disk where the operating system resides.
- When multiple mirror disk resources are used, it is recommended to prepare (adding) a disk per mirror disk resource. Allocating multiple mirror disk resources on the same disk may result in degraded performance and it may take a while to complete mirror recovery due to disk access performance on Linux operating system.
- Disks used for mirroring must be the same in all servers.
- Disk interface

  Mirror disks on both servers and disks where mirror partition is allocated should be of the same disk interface

  Example

Combination	server1	server2
OK	SCSI	SCSI
OK	IDE	IDE
NG	IDE	SCSI

- Disk type

  Mirror disks on both servers and disks where mirror partition is allocated should be of the same disk type

  Example

Combination	server1	server2
OK	HDD	HDD
OK	SSD	SSD
NG	HDD	SSD

- Sector size

  Mirror disks on both servers and disks where mirror partition is allocated should be of the same sector size

  Example

Combination	server1	server2
OK	512B	512B
OK	4KB	4KB

Continued on next page

Table 6.4 – continued from previous page

Combination	server1	server2
NG	512B	4KB

- Notes when the geometries of the disks used as mirror disks differ between the servers.

  The partition size allocated by the fdisk command is aligned by the number of blocks (units) per cylinder. Allocate a data partition considering the relationship between data partition size and direction for initial mirror configuration to be as indicated below:

  Source server <= Destination server

  "Source server" refers to the server where the failover group that a mirror disk resource belongs has a higher priority in failover policy. "Destination server" refers to the server where the failover group that a mirror disk resource belongs has a lower priority in failover policy.

  Make sure that the data partition sizes do not cross over 32GiB, 64GiB, 96GiB, and so on (multiples of 32GiB) on the source server and the destination server. For sizes that cross over multiples of 32GiB, initial mirror construction may fail. Be careful, therefore, to secure data partitions of similar sizes.

  Example

Combination	Data partition size		Description
	On server 1	On server 2	
OK	30GiB	31GiB	OK because both are in the range of 0 to 32GiB.
OK	50GiB	60GiB	OK because both are in the range of 32GiB to 64GiB.
NG	30GiB	39GiB	Error because they are crossing over 32GiB.
NG	60GiB	70GiB	Error because they are crossing over 64GiB.

6.1.3 Hardware requirements for shared disks

- When a Linux LVM stripe set, volume set, mirroring, or stripe set with parity is used:
- EXPRESSCLUSTER cannot control ReadOnly/ReadWrite of the partition configured for the disk resource.
- When you use LVM features, use the disk resource (disk type: "lvm") and the volume manager resource.

The example of disk configuration is shown below.
More than one LUNs are stored in the actual disk. In the following figure, there are disk heartbeat-dedicated LUN and disk groups, dg1 and dg2 clustering multiple disks.
Moreover, the volumes, vxvol1, vxvol2 are allocated in dg1 and the volumes, vxvol3, vxvol4 are allocated in dg2. The volume comprises partitions allocated in disk groups.
In EXPRESSCLUSTER, a disk group is defined as a VxVM disk group resource, and a volume is defined as a VxVM volume resource.

6.1.4 Hardware requirements for hybrid disks

- Disks to be used as a hybrid disk resource do not support a Linux md stripe set, volume set, mirroring, and stripe set with parity.
- Linux LVM volumes can be used for both cluster partitions and data partitions.
  For SuSE, however, LVM and MultiPath volumes cannot be used for data partitions. (This is because for SuSE, ReadOnly or ReadWrite control over these volumes cannot be performed by EXPRESSCLUSTER.)
- Hybrid disk resource cannot be made as a target of a Linux md stripe set, volume set, mirroring, and stripe set with parity.
- Hybrid partitions (data partition and cluster partition) are required to use a hybrid disk resource.
- When a disk for hybrid disk is allocated in the shared disk, a partition for disk heartbeat resource between servers sharing the shared disk device is required.
- The following are the two ways to allocate partitions when a disk for hybrid disk is allocated from a disk which is not a shared disk:
  - Allocate hybrid partitions (data partition and cluster partition) on the disk where the operating system (such as root partition and swap partition) resides.
  - Reserve (or add) a disk (or LUN) not used by the operating system and allocate a hybrid partition on the disk.
- Consider the following when allocating hybrid partitions:
  - When maintainability and performance are important:
    - It is recommended to have a hybrid disk that is not used by the OS.
  - When LUN cannot be added due to hardware RAID specification or when changing LUN configuration is difficult in hardware RAID pre-install model:
    - Allocate a hybrid partition on the same disk where the operating system resides.

	Device for which hybrid disk resource is allocated	
Type of required partition	Shared disk device	Non-shared disk device
Data partition	Required	Required
Cluster partition	Required	Required
Partition for disk heart beat	Required	Not Required
Allocation on the same disk (LUN) as where the OS is	-	Possible

- When multiple hybrid disk resources are used, it is recommended to prepare (add) a LUN per hybrid disk resource. Allocating multiple hybrid disk resources on the same disk may result in degraded in performance and it may take a while to complete mirror recovery due to disk access performance on Linux operating system.
- Notes when the geometries of the disks used as hybrid disks differ between the servers.

  Allocate a data partition considering the relationship between data partition size and direction for initial mirror configuration to be as indicated below:

  Source server <= Destination server

  "Source server" refers to the server with a higher priority in failover policy in the failover group where the hybrid disk resource belongs. "Destination server" refers to the server with a lower priority in failover policy in the failover group where the hybrid disk resource belongs has.

Make sure that the data partition sizes do not cross over 32GiB, 64GiB, 96GiB, and so on (multiples of 32GiB) on the source server and the destination server. For sizes that cross over multiples of 32GiB, initial mirror construction may fail. Be careful, therefore, to secure data partitions of similar sizes.

Example

Combination	Data partition size		Description
	On server 1	On server 2	
OK	30GiB	31GiB	OK because both are in the range of 0 to 32GiB.
OK	50GiB	60GiB	OK because both are in the range of 32GiB to 64GiB.
NG	30GiB	39GiB	Error because they are crossing over 32GiB.
NG	60GiB	70GiB	Error because they are crossing over 64GiB.

6.1.5 IPv6 environment

The following function cannot be used in an IPv6 environment:

- AWS Elastic IP resource
- AWS Virtual IP resource
- AWS Secondary IP resource
- AWS DNS resource
- Azure probe port resource
- Azure DNS resource
- Google Cloud virtual IP resource
- Google Cloud DNS resource
- Oracle Cloud virtual IP resource
- AWS Elastic IP monitor
- AWS Virtual IP monitor
- AWS Secondary IP monitor
- AWS AZ monitor
- AWS DNS monitor
- Azure probe port monitor
- Azure load balance monitor
- Azure DNS monitor
- Google Cloud virtual IP monitor resource
- Google Cloud load balance monitor resource
- Google Cloud DNS monitor resource
- Oracle Cloud virtual IP monitor resource
- Oracle Cloud load balance monitor resource

The following functions cannot use link-local addresses:

- LAN heartbeat resource
- Kernel mode LAN heartbeat resource
- Mirror disk connect
- PING network partition resolution resource
- FIP resource
- VIP resource

6.1.6 Network configuration

The cluster configuration cannot be configured or operated in an environment, such as NAT, where an IP address of a local server is different from that of a remote server.

Example of network configuration

The following figure shows two servers connected to different networks with a NAT device set between them.
For example, assume that the NAT device is set as "the packet from the external network to 10.0.0.2 is forwarded to the internal network."
However, to build a cluster with Server 1 and Server 2 in this environment, IP addresses for different networks must be specified in each server.
In the environment with each server set in different subnets like this, a cluster cannot be properly configured or operated.

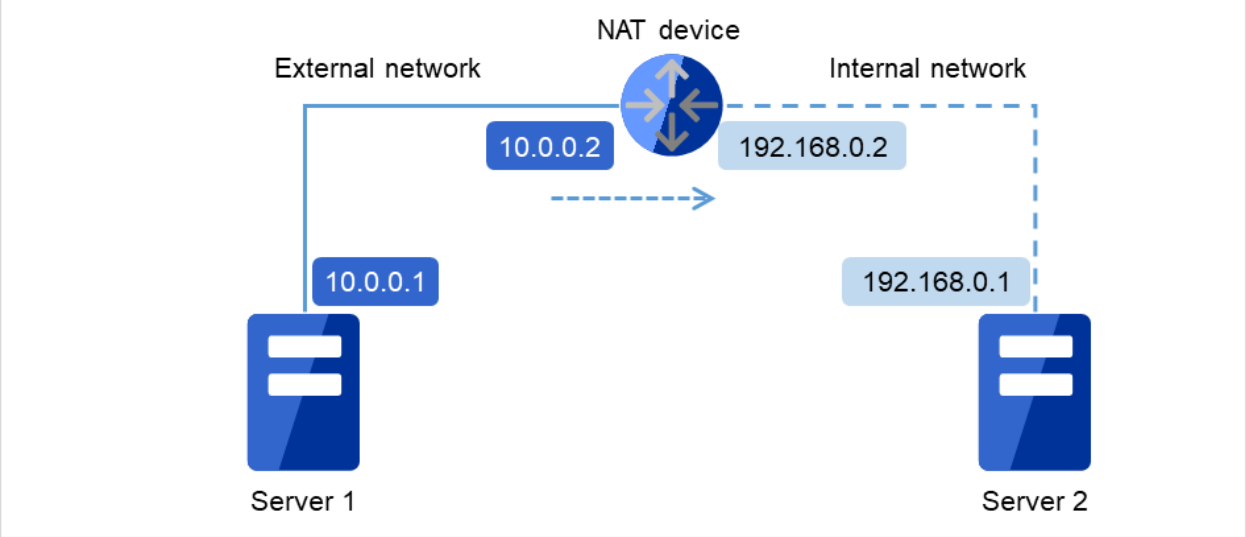


Fig. 6.1: Example of the environment where a cluster cannot be configured

- Cluster settings for Server 1
  - Local server: 10.0.0.1
  - Remote server: 10.0.0.2

- Cluster settings for Server 2
  - Local server: 192.168.0.1
  - Remote server: 10.0.0.1

6.1.7 Execute Script before Final Action setting for monitor resource recovery action

EXPRESSCLUSTER version 3.1.0-1 and later supports the execution of a script before reactivation and before failover.
The same script is executed in either case. Therefore, if **Execute Script before Final Action** is set with a version earlier than 3.1.0-1, editing of the script file may be required.
For the additional script configuration needed to execute the script before reactivation and before failover, the script file must be edited to assign processing to each recovery action.
For the assignment of processing for a recovery action, see "Recovery/pre-recovery action script" in "Monitor resource details" in the "Reference Guide".

6.1.8 NIC Link Up/Down monitor resource

Some NIC boards and drivers do not support required ioctl( ).
The propriety of a NIC Link Up/Down monitor resource of operation can be checked by the ethtool command which each distributor offers.

```
ethtool eth0
Settings for eth0:
    Supported ports: [ TP ]
    Supported link modes:  10baseT/Half 10baseT/Full
                           100baseT/Half 100baseT/Full
                           1000baseT/Full
    Supports auto-negotiation: Yes
    Advertised link modes: 10baseT/Half 10baseT/Full
                           100baseT/Half 100baseT/Full
                           1000baseT/Full
    Advertised auto-negotiation: Yes
    Speed: 1000Mb/s
    Duplex: Full
    Port: Twisted Pair
    PHYAD: 0
    Transceiver: internal
    Auto-negotiation: on
    Supports Wake-on: umbg
    Wake-on: g
    Current message level: 0x00000007 (7)
    Link detected: yes
```

- When the LAN cable link status ("Link detected: yes") is not displayed as the result of the ethtool command:
  - It is highly likely that NIC Link Up/Down monitor resource of EXPRESSCLUSTER is not operable. Use IP monitor resource instead.
- When the LAN cable link status ("Link detected: yes") is displayed as the result of the ethtool command:

- In most cases NIC Link Up/Down monitor resource of EXPRESSCLUSTER can be operated, but sometimes it cannot be operated.
- Particularly in the following hardware, NIC Link Up/Down monitor resource of EXPRESSCLUSTER may not be operated. Use IP monitor resource instead.
- When hardware is installed between the actual LAN connector and NIC chip such as a blade server
- When the monitored NIC is in a bonding environment, check whether the MII Polling Interval is set to 0 or higher.

To check if NIC Link Up/Down monitor resource can be used by using EXPRESSCLUSTER on an actual machine, follow the steps below to check the operation.

1. Register NIC Link Up/Down monitor resource with the configuration information.
   Select **No Operation** for the configuration of recovery operation of NIC Link Up/Down monitor resource upon failure detection.

1. Start the cluster.
2. Check the status of NIC Link Up/Down monitor resource.
   If the status of NIC Link Up/Down monitor resource is abnormal while LAN cable link status is normal, NIC Link Up/Down monitor resource cannot be operated.
3. If NIC Link Up/Down monitor resource status becomes abnormal when LAN cable link status is made abnormal status (link down status), NIC Link Up/Down monitor resource can be operated.
   If the status remains to be normal, NIC Link Up/Down monitor resource cannot be operated.

6.1.9 Write function of the Mirror disk resource and Hybrid disk resource

There are 2 types of disk mirroring of mirror disk resources and hybrid disk resources: synchronous mirroring and asynchronous mirroring.

In synchronous mirroring, data is written in the disks of both servers for every request to write data in the data partition to be mirrored and its completion is waited. Data is written in each of the servers along with this, but it is written in disks of other servers via network, so writing performance declines more significantly compared to a normal local disk that is not to be mirrored. In case of the remote cluster configuration, since the network communication speed is slow and delay is long, the writing performance declines drastically.

In asynchronous mirroring, data is written to the local server immediately. However, when writing data to other server, it is saved to the local queue first and then written in the background. In case of asynchronous mirror, the data to be updated is saved in the queue for every writing request as well, so the writing performance declines more significantly, compared to the normal local disk that is not to be mirrored and the shared disk. For this reason, it is recommended to use the shared disk for the system (such as the database system with lots of update systems) that is required high throughput for writing data in disks.

In case of asynchronous mirroring, the writing sequence will be guaranteed, but the data that has been updated to the latest may be lost, if an active server shuts down. For this reason, if it is required to inherit the data immediately before an error occurs for sure, use synchronous mirroring or the shared disk.

6.1.10 History file of asynchronous mirroring

In mirror disk or hybrid disk with asynchronous mode, data that cannot afford to be written in memory queue is recorded temporarily in a folder specified to save history files. When the limit of the file is not specified, history files are written in the specified folder without limitation. In this case, the line speed is too low, compared to the disk update amount of application, writing data to other server cannot catch up with update of data on the disk, and history files will overflow from the disk.

For this reason, it is required to reserve a communication line with enough speed in the remote cluster configuration as well, in accordance with the disk update amount of application.

It is also required to prepare against an overflow of history files from the directory due to a long delay in communication or a continuous update of data on the disk. The preparation can be achieved by maintaining enough free space and limiting the history file size for the history file directory, or by specifying a different directory on a non-system disk.

6.1.11 Not outputting syslog to the Mirror disk resource or the Hybrid disk resource

Do not set directories or subdirectories which mounted the mirror disk resource or the hybrid disk resource as syslog output destination directories.
When the mirror disk connection is disconnected, the I/O to the mirror partition may stop until the disconnection is detected. The system may become abnormal because of the syslog output stoppage at this time.
When outputting syslog to the mirror disk resource or the hybrid disk resource is necessary, consider the followings.

- Use bonding as a way of path redundancy of the mirror disk connection.
- Adjust the user-mode monitoring timeout value or the mirror related timeout values.

6.1.12 Notes when terminating the Mirror disk resource or the Hybrid disk resource

- In case that processes which access to the directories, subdirectories and files which mounted the mirror disk resource or the hybrid disk resource exist, terminate the accesses to each disk resource by using ending script or other methods at deactivation of each disk resource like when shutdown or failover.

  Depending on the settings of each disk resource, action at abnormity detection when unmounting (forcibly terminate processes while each disk resource is being accessed) may occur, or recovery action at deactivation failure caused by unmount failure (OS shutdown or other actions) may be executed.
- In case that a massive amount of accesses to directories, subdirectories or files which mounted the mirror disk resource or hybrid disk resource are executed, it may take much time before the cache of the file systems is written out to the disks when unmounting at disk resource deactivation.

  At times like this, set the timeout interval of unmount longer enough so that the writing to the disks will successfully complete.
- For the details of this setting,

  see "Group resource details" in "Reference Guide",

  Recovery Operation tab or **Mirror Disk Resource Tuning Properties** or **Unmount tab** in **Details tab** in "Understanding Mirror disk resources" or "Understanding Hybrid disk resources".

6.1.13 Data consistency among multiple asynchronous mirror disks

In mirror disk or hybrid disk with asynchronous mode, writing data to the data partition of the active server is performed in the same order as the data partition of the standby server.

This writing order is guaranteed except during the initial mirror disk configuration or recovery (copy) period after suspending mirroring the disks. The data consistency among the files on the standby data partition is guaranteed.

However, the writing order is not guaranteed among multiple mirror disk resources and hybrid disk resources. For example, if a file gets older than the other and files that cannot maintain the data consistency are distributed to multiple asynchronous mirror disks, an application may not run properly when it fails over due to server failure.

For this reason, be sure to place these files on the same asynchronous mirror disk or hybrid disk.

6.1.14 Mirror data reference at the synchronization destination if mirror synchronization is interrupted

If mirror synchronization is interrupted for a mirror disk or a hybrid disk in the mirror synchronization state, using the mirror disks or the clpmdctrl / clphdctrl command (with the --break / -b / --nosync option specified), the file system and application data may be abnormal if the mirror disk on the server on the mirror synchronization destination (copy destination) is made accessible by performing forced activation (removing the access restriction) or forced mirror recovery.

This occurs because if mirror synchronization is interrupted on the server on the mirror synchronization source (server on which the resources are activated) leading to an inconsistent state in which there are portions that can be synchronized with the synchronization destination and portions that cannot be synchronized such as; for example, when an application is writing to a mirror disk area, part of the data and so on will be retained in the cache and so on (memory) of the OS, but not yet actually written to the mirror disk, or may be in the process of being written.

If you want to perform access in a state in which consistency with the mirror disk on the mirror synchronization destination (standby server) is ensured, secure a rest point on the mirror synchronization source (active server on which the resources are activated) first and then interrupt mirror synchronization. Alternatively, secure a rest point by deactivating. (With an application end, access to the mirror area ends, and by unmounting the mirror disk, the cache and so on of the OS are all written to the mirror disk.)

Similarly, if mirror recovery is interrupted for a mirror disk or a hybrid disk that is in the middle of mirror recovery (mirror resynchronization), the file system and application data may be abnormal if the mirror disk on the mirror synchronization destination is accessed by performing forced activation (removing the access restriction) or forced mirror recovery.

This also occurs because mirror recovery is interrupted in an inconsistent state in which there are portions that can be synchronized but also portions that cannot.

6.1.15 O_DIRECT for mirror or hybrid disk resources

Do not use the O_DIRECT flag of the open() system call for the mirror partition device (/dev/NMPx) of a mirror or hybrid disk resource.

Examples include the Oracle parameter filesystemio_options = setall.

Do not specify the O_DIRECT mode of the disk monitor resource for the mirror partition device (/dev/NMPx) of a mirror or hybrid disk resource.

6.1.16 Initial mirror construction time for mirror or hybrid disk resources

The time that takes to construct the initial mirror is different between ext2/ext3/ext4/xfs and other file systems.

Note: xfs shortens the time with resource deactivation.

6.1.17 Mirror or hybrid disk connect

- When using redundant mirror or hybrid disk connect, both version of IP address are needed to be the same.
- All the IP addresses used by mirror disk connect must be set to IPv4 or IPv6.

6.1.18 JVM monitor resources

- Up to 25 Java VMs can be monitored concurrently. The Java VMs that can be monitored concurrently are those which are uniquely identified by the Cluster WebUI (with **Identifier** in the **Monitor (special)** tab).
- Connections between Java VMs and Java Resource Agent do not support SSL.
- It may not be possible to detect thread deadlocks. This is a known problem in Java VM. For details, refer to "Bug ID: 6380127" in the Oracle Bug Database.
- The JVM monitor resources can monitor only the Java VMs on the server on which the JVM monitor resources are running.
- The JVM monitor resources can monitor only one JBoss server instance per server.
- The Java installation path setting made by the Cluster WebUI (with **Java Installation Path** in the **JVM monitor** tab in **Cluster Properties**) is shared by the servers in the cluster. The version and update of Java VM used for JVM monitoring must be the same on every server in the cluster.
- The management port number setting made by the Cluster WebUI (with **Management Port** in the **Connection Setting** dialog box opened from the **JVM monitor** tab in **Cluster Properties**) is shared by all the servers in the cluster.
- Application monitoring is disabled when an application to be monitored on the IA32 version is running on an x86_64 version OS.
- If a large value such as 3,000 or more is specified as the maximum Java heap size by the Cluster WebUI (by using Maximum Java Heap Size on the **JVM monitor** tab in **Cluster Properties**), The JVM monitor resources will fail to start up. The maximum heap size differs depending on the environment, so be sure to specify a value based on the capacity of the mounted system memory.
- If "-XX:+UseG1GC" is added as a startup option of the target Java VM, the settings on the **Memory** tab on the **Monitor(special)** tab in **Properties** of JVM monitor resources cannot be monitored before Java 7.
  It's possible to monitor by choosing **Oracle Java (usage monitoring)** in **JVM Type** on the **Monitor(special)** tab after Java 8.

6.1.19 Mail reporting

The mail reporting function is not supported by STARTTLS and SSL.

6.1.20 Requirements for network warning light

- When using "DN-1000S" or "DN-1500GL," do not set your password for the warning light.
- To play an audio file as a warning, you must register the audio file to a network warning light supporting audio file playback.

  For details about how to register an audio file, see the manual of the network warning light you want to use.
- Set up a network warning light so that a server in a cluster is permitted to execute the rsh command to that warning light.

6.2 Installing operating system

Notes on parameters to be determined when installing an operating system, allocating resources, and naming rules are described in this section.

6.2.1 Mirror disks

- Disk partition

  Example: When adding one SCSI disk to each of both servers and making a pair of mirrored disks:

  In the figure below, a SCSI disk is added to each of two servers.
  The inside of the disk is divided into the cluster partition and the data partition. This set of partitions, called a mirror partition device, is a unit for the failover of the mirror disk resource.

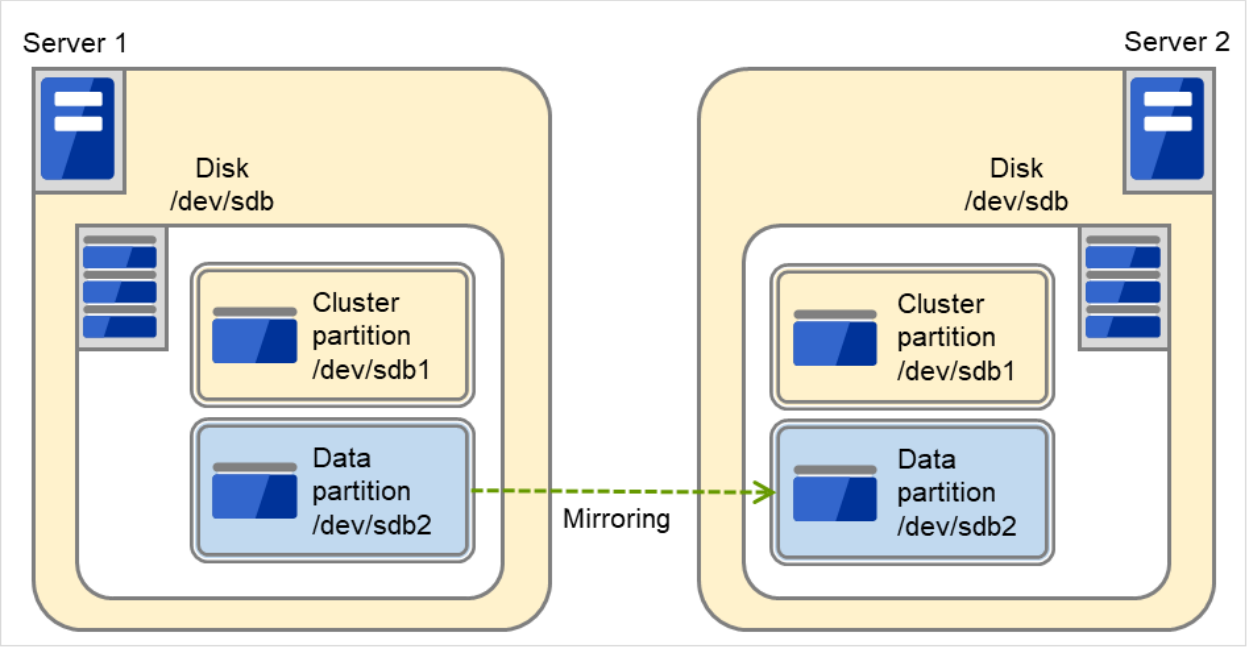


Fig. 6.2: Configuration of disks and partitions (with SCSI disks added)

Example: When using free space of IDE disks of both servers, where the OS is stored, and making a pair of mirrored disks:

The following figure illustrates using the free space of each built-in disk as a mirror partition device (cluster partition and data partition):

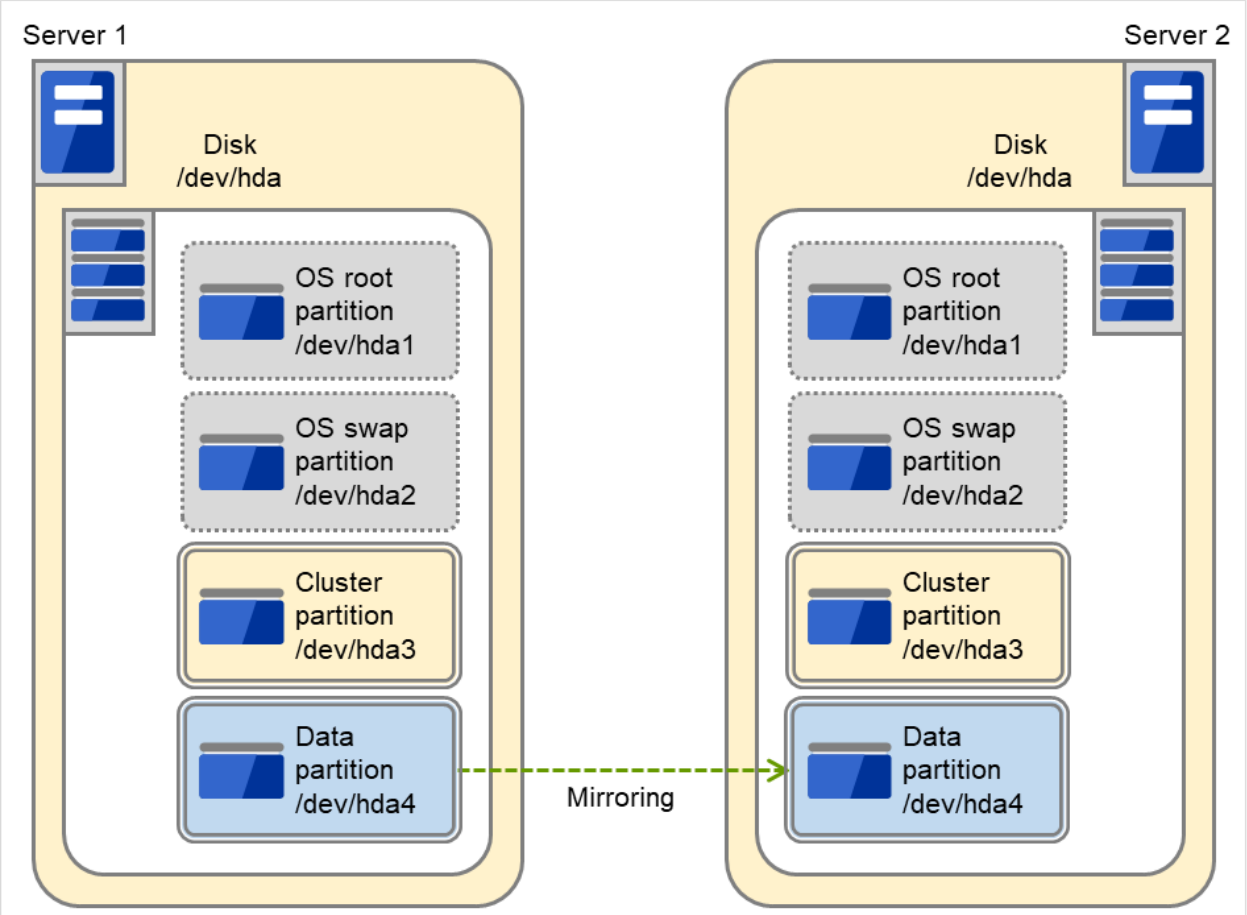


Fig. 6.3: Configuration of disks and partitions (with free space of existing disks used)

- Mirror partition device refers to cluster partition and data partition.
- Allocate cluster partition and data partition on each server as a pair.
- It is possible to allocate a mirror partition (cluster partition and data partition) on the disk where the operating system resides (such as root partition and swap partition.).
  - When maintainability and performance are important:
    It is recommended to have a mirror disk that is not used by the operating system (such as root partition and swap partition.)
  - When LUN cannot be added due to hardware RAID specification: or
    When changing LUN configuration is difficult in hardware RAID pre-install model:

  It is possible to allocate a mirror partition (cluster partition and data partition) on the disk where the operating system resides (such as root partition and swap partition.)

- Disk configurations

  Multiple disks can be used as mirror disks on a single server. Or, you can allocate multiple mirror partitions on a single disk.

  Example: When adding two SCSI disks to each of both servers and making two pairs of mirrored disks:

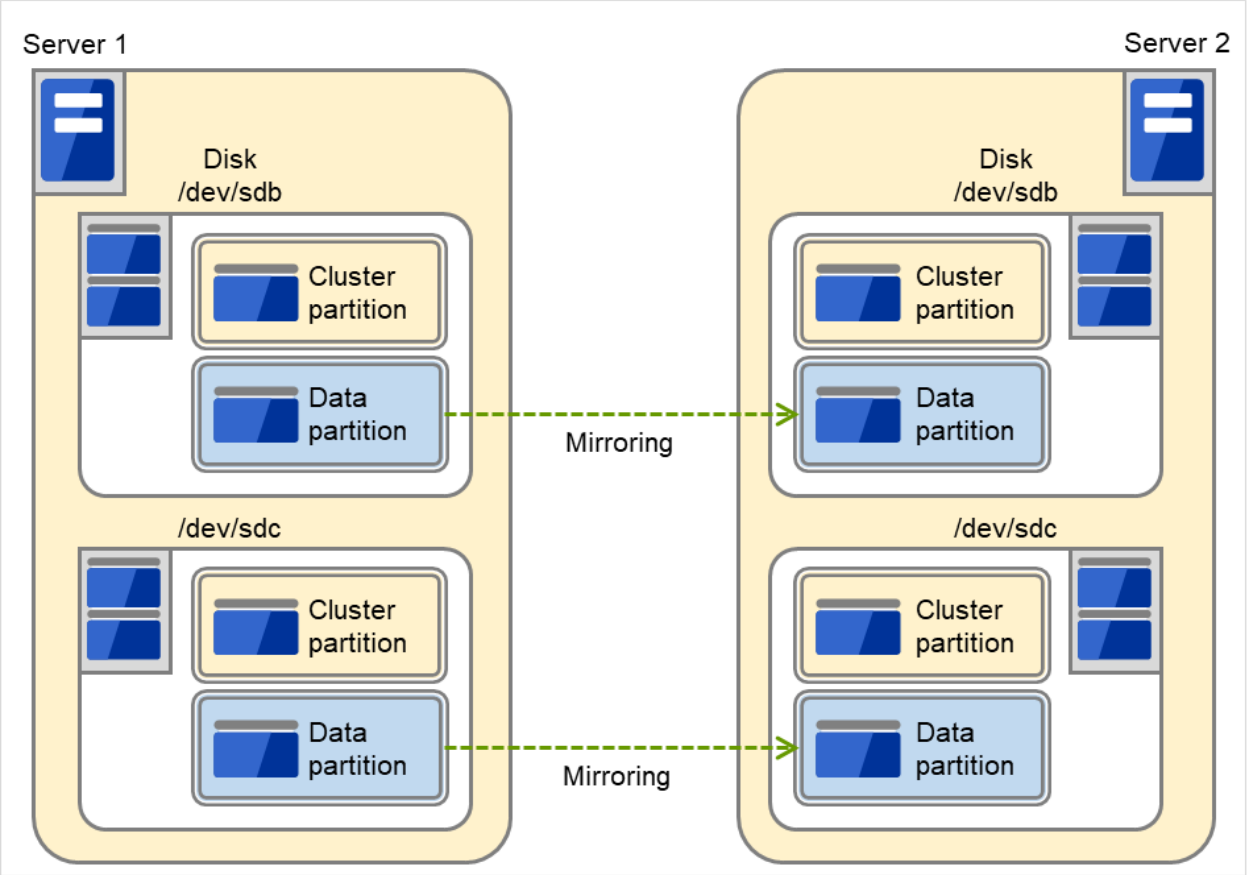


Fig. 6.4: Each of multiple disks used as a mirror partition

- Allocate two partitions, cluster partition and data partition, as a pair on each disk.
- Use of the data partition as the first disk and the cluster partition as the second disk is not permitted.

Example: When adding one SCSI disk to each of both servers and making two mirror partitions:

The figure below illustrates the case where two mirror partitions are allocated in a disk.

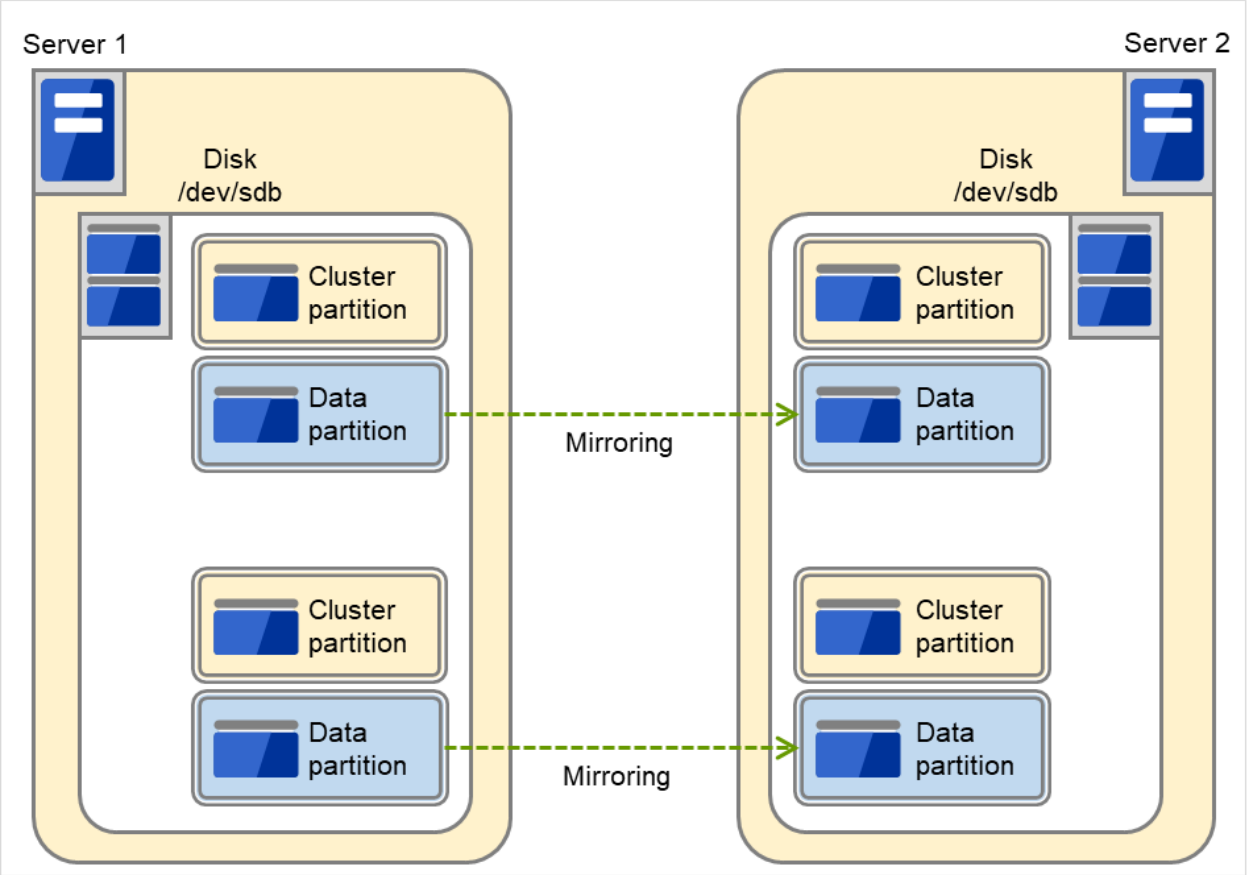


Fig. 6.5: Multiple areas of each disk used as mirror partitions

– A disk does not support a Linux md stripe set, volume set, mirroring, and stripe set with parity.

6.2.2 Hybrid disks

- Disk partition

  Disks that are shared or not shared (server with built-in disk, external disk chassis not shared by servers etc.) can be used.

Example) When two servers use a shared disk and the third server uses a built-in disk in the server:

In the figure below, the built-in disk of Server 3 is used as a mirror partition device.

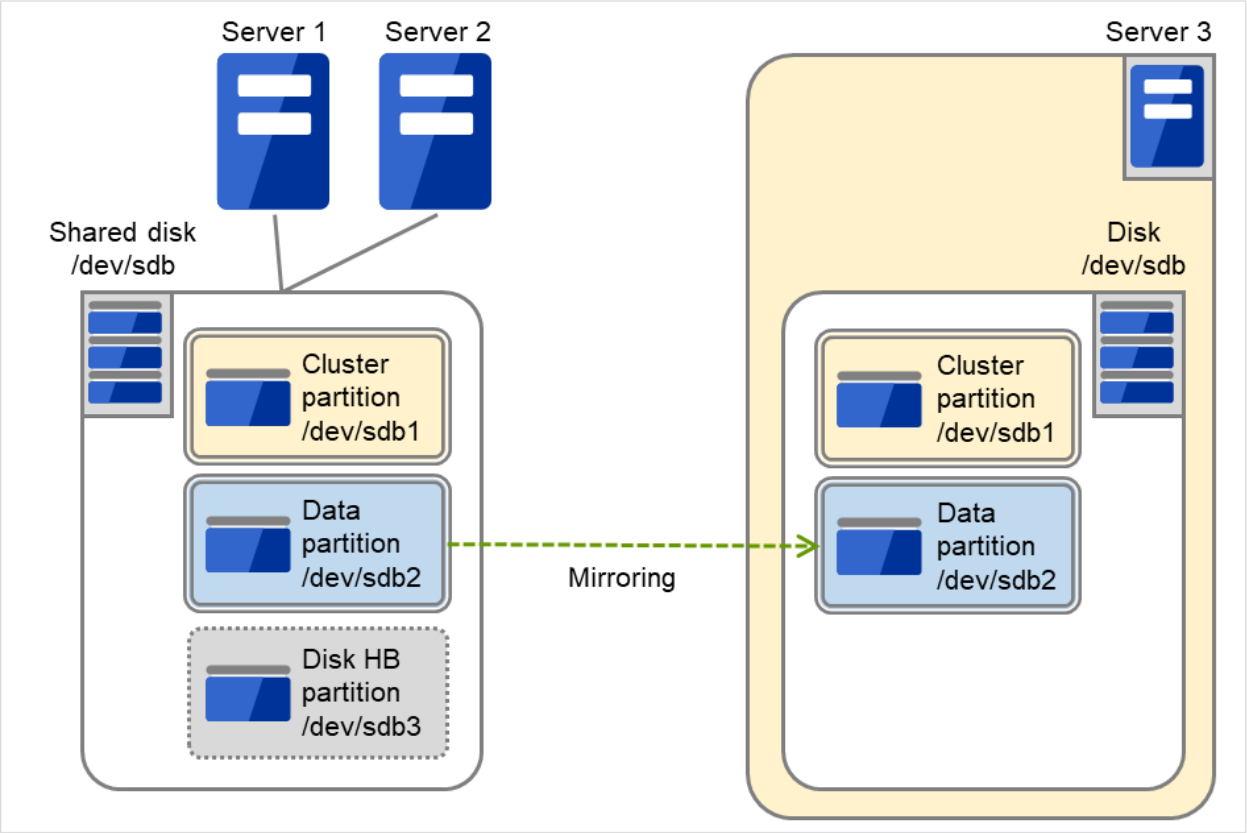


Fig. 6.6: Configuration of disks and partitions (with a shared disk and a built-in disk used)

- Mirror partition device is a device EXPRESSCLUSTER mirroring driver provides in the upper.
- Allocate cluster partition and data partition on each server as a pair.
- When a disk that is not shared (e.g. server with a built-in disk, external disk chassis that is not shared among servers) is used, it is possible to allocate mirror partitions (cluster partition and data partition) on the disk where the operating system resides (such as root partition and swap partition.).
  - When maintainability and performance are important:
    It is recommended to have a mirror disk that is not used by the operating system (such as root partition and swap partition.)
  - When LUN cannot be added due to hardware RAID specification: or
    When changing LUN configuration is difficult in hardware RAID pre-install model:

    It is possible to allocate mirror partitions (cluster partition and data partition) on the disk where the operating system resides (such as root partition and swap partition.)

- When a hybrid disk is allocated in a shared disk device, allocate a partition for the disk heart beat resource between servers sharing the shared disk device.
- A disk does not support a Linux md stripe set, volume set, mirroring, and stripe set with parity.

6.2.3 Dependent library

- libxml2

  Install libxml2 when installing the operating system.

6.2.4 Dependent driver

- softdog

  This driver is necessary when softdog is used to monitor user-mode monitor resource.

  Configure a loadable module. Static driver cannot be used.

6.2.5 Necessary package

When you install the OS, install the following packages as well:

- tar
- NetworkManager-config-server

6.2.6 The major number of Mirror driver

Use mirror driver's major number 218. Do not use major number 218 for other device drivers.

6.2.7 The major number of Kernel mode LAN heartbeat and keepalive drivers

- Use major number 10, minor number 253 for kernel mode LAN heartbeat driver.
- Use major number 10, minor number 254 for keepalive driver.

Make sure to check that other drivers are not using major and minor numbers described above.

6.2.8 Partition for RAW monitoring of disk monitor resources

- Allocate a partition for monitoring when setting up RAW monitoring of disk monitor resources. The partition size should be 10MB.

6.2.9 SELinux settings

- If you want to load the drivers of EXPRESSCLUSTER with enforcing specified for SELinux, complete the procedure as described in the following: "Installation and Configuration Guide" -> "SELinux settings (Required)"

6.2.10 NetworkManager settings

- If the NetworkManager service is running in a Red Hat Enterprise Linux 6 environment, an unintended behavior (such as detouring the communication path, or disappearance of the network interface) may occur upon disconnection of the network. It is recommended to set NetworkManager to stop the service.

6.2.11 LVM metadata daemon settings

- When controlling or monitoring the LVM by using the volume manager resource or volume manager monitor resource in an environment of Red Hat Enterprise Linux 7 or later, the LVM metadata daemon must be disabled.

  The procedure to disable the metadata daemon is as follows:

  (1) Execute the following command to stop the LVM metadata daemon.

  ```
  # systemctl stop lvm2-lvmetad.service
  ```

  (2) Edit /etc/lvm/lvm.conf to set the value of use_lvmetad to 0.

6.2.12 Secure Boot settings

- Disable the Secure Boot settings.

6.3 Before installing EXPRESSCLUSTER

Notes after installing an operating system, when configuring OS and disks are described in this section.

6.3.1 Communication port number

In EXPRESSCLUSTER, the following port numbers are used. You can change the port number by using the Cluster WebUI.
Make sure not to access the following port numbers from a program other than EXPRESSCLUSTER.

Configure to be able to access the port number below when setting a firewall on a server.

After installing EXPRESSCLUSTER, you can use the clpfwctrl command to configure a firewall. For more information, see "Reference Guide" -> "EXPRESSCLUSTER command reference" -> "Adding a firewall rule (clpfwctrl command)". Ports to be set with the clpfwctrl command are marked with ✓ in the clpfwctrl column of the table below.

For a cloud environment, allow access to ports numbered as below, not only in a firewall configuration at the instance side but also in a security configuration at the cloud infrastructure side.

- **Server to Server Loopback in servers**

From		To		Used for	clpfwctrl
Server	Automatic allocation[6]	Server	29001/TCP	Internal communication	✓
Server	Automatic allocation	Server	29002/TCP	Data transfer	✓
Server	Automatic allocation	Server	29002/UDP	Heartbeat	✓
Server	Automatic allocation	Server	29003/UDP	Alert synchronization	✓
Server	Automatic allocation	Server	29004/TCP	Communication between mirror agents	✓
Server	Automatic allocation	Server	29006/UDP	Heartbeat (kernel mode)	✓
Server	Automatic allocation	Server	29008/TCP	Cluster information management	✓
Server	Automatic allocation	Server	29010/TCP	Internal communication of RESTful API	✓
Server	Automatic allocation	Server	XXXX[7] /TCP	Mirror disk resource data synchronization	✓
Server	Automatic allocation	Server	XXXX[8] /TCP	Communication between mirror drivers	✓
Server	Automatic allocation	Server	XXXX[9] /TCP	Communication between mirror drivers	✓

Continued on next page

Table 6.5 – continued from previous page

From		To		Used for	clpfwctrl
Server	icmp	Server	icmp	keepalive between mirror drivers, duplication check for FIP/VIP resource and mirror agent	
Server	Automatic allocation	Server	XXXX[10] /UDP	Internal log communication	✓

- **Client to Server**

From		To		Used for	clpfwctrl
RESTful API client	Automatic allocation	Server	29009/TCP	http communication	✓

- **Cluster WebUI to Server**

From		To		Used for	clpfwctrl
Cluster WebUI	Automatic allocation	Server	29003/TCP	http communication	✓

- **Others**

From		To		Used for	clpfwctrl
Server	Automatic allocation	Network warning light	See the manual for each product.	Network warning light control	
Server	Automatic allocation	Management LAN of server BMC	623/UDP	BMC control (Forced stop)	
Server	Automatic allocation	Witness server	Communication port number specified with Cluster WebUI	Connection destination host of the Witness heartbeat resource	
Server	icmp	Monitoring target	icmp	IP monitor	

Continued on next page

[6] In automatic allocation, a port number not being used at a given time is allocated.

[7] This is a port number used per mirror disk resource or hybrid disk resource and is set when creating mirror disk resource or hybrid disk resource. A port number 29051 is set by default. When you add a mirror disk resource or hybrid disk resource, this value is automatically incremented by 1. To change the value, click **Details** tab in the **[md] Resource Properties** or the **[hd] Resource Properties** dialog box of the Cluster WebUI. For more information, refer to "Group resource details" in the "Reference Guide".

[8] This is a port number used per mirror disk resource or hybrid disk resource and is set when creating mirror disk resource or hybrid disk resource. A port number 29031 is set by default. When you add a mirror disk resource or a hybrid disk resource, this value is automatically incremented by 1. To change the value, click **Details** tab in the **[md] Resource Properties** or the **[hd] Resource Properties** dialog box of the Cluster WebUI. For more information, refer to "Group resource details" in the "Reference Guide".

[9] This is a port number used per mirror disk resource or hybrid disk resource and is set when creating mirror disk resource or hybrid disk resource. A port number 29071 is set by default. When you add a mirror disk resource or hybrid disk resource this value is automatically incremented by 1. To change the value, click **Details** tab in the **[md] Resource Properties** or the **[hd] Resource Properties** dialog box of the Cluster WebUI. For more information, refer to "Group resource details" in the "Reference Guide".

[10] Select **UDP** for the **Communication Method for Internal Logs** in the **Port No. (Log)** tab in **Cluster Properties**. Use the port number configured in Port No. Communication port is not used for the default log communication method **UNIX Domain**.

Table 6.8 – continued from previous page

From		To		Used for	clpfwctrl
Server	icmp	Monitoring target	icmp	Monitoring target of Ping method network partition resolution resource	
Server	Automatic allocation	Monitoring target	Management port number set by the Cluster WebUI	Monitoring target of HTTP method of network partition resolution resource	
Server	Automatic allocation	Server	Management port number set by the Cluster WebUI[11]	JVM monitor	✓
Server	Automatic allocation	Monitoring target	Connection port number set by the Cluster WebUI[11]	JVM monitor	
Server	Automatic allocation	Server	Probe port number set by the Cluster WebUI[12]	Azure probe port resource	✓

Continued on next page

Table 6.8 – continued from previous page

From		To		Used for	clpfwctrl
Server	Automatic allocation	AWS region endpoint	443/tcp[13]	AWS elastic ip resource AWS virtual ip resource AWS secondary ip resource AWS DNS resource AWS elastic ip monitor resource AWS virtual ip monitor resource AWS secondary ip monitor resource AWS AZ monitor resource AWS DNS monitor resource AWS forced stop resource	
Server	Automatic allocation	Azure endpoint	443/tcp[14]	Azure DNS resource	
Server	Automatic allocation	Azure authoritative name server	53/udp	Azure DNS monitor resource	
Server	Automatic allocation	Server	Port number set in Cluster WebUI[12]	Google Cloud virtual IP resource	✓
Server	Automatic allocation	Server	Port number set in Cluster WebUI[12]	Oracle Cloud virtual IP resource	✓

[11] The JVM monitor resource uses the following two port numbers.
- A management port number is a port number that the JVM monitor resource internally uses. To set this number, use the **Connection Setting** dialog box opened from the **JVM monitor** tab in **Cluster Properties** of the Cluster WebUI. For details, refer to "Parameter details" in the "Reference Guide".
- A connection port number is used to establish a connection to the target Java VM (WebLogic Server or WebOTX). To set this number, use the **Monitor (special)** tab in **Properties** of the Cluster WebUI for the corresponding JVM monitor resource. For details, refer to "Monitor resource details" in the "Reference Guide".

[12] Port number used by the load balancer for the alive monitoring of each server.
[13] The above port numbers are used with the AWS CLI, which is executed by the following AWS-related resources:
[14] The Azure DNS resource runs the Azure CLI. The above port numbers are used by the Azure CLI.

- AWS Elastic IP resource
- AWS Virtual IP resource
- AWS Secondary IP resource
- AWS DNS resource
- AWS Elastic IP monitor resource
- AWS Virtual IP monitor resource
- AWS Secondary IP monitor resource
- AWS AZ monitor resource
- AWS DNS monitor resource
- AWS Forced stop resource

6.3.2 Changing the range of automatic allocation for the communication port numbers

- The range of automatic allocation for the communication port numbers managed by the OS might overlap the communication port numbers used by EXPRESSCLUSTER.
- Change the OS settings to avoid duplication when the range of automatic allocation for the communication numbers managed by OS and the communication numbers used by EXPRESSCLUSTER are duplicated.

Examples of checking and displaying OS setting conditions.

The range of automatic allocation for the communication port numbers depends on the distribution.

```
# cat /proc/sys/net/ipv4/ip_local_port_range
1024 65000
```

This is the condition to be assigned for the range from 1024 to 65000 when the application requests automatic allocation for the communication port numbers to the OS.

```
# cat /proc/sys/net/ipv4/ip_local_port_range
32768 61000
```

This is the condition to be assigned for the range from 32768 to 61000 when the application requests automatic allocation for the communication port numbers to the OS.

Examples of OS settings change

Add the line below to /etc/sysctl.conf. (When changing to the range from 30000 to 65000)

```
net.ipv4.ip_local_port_range = 30000 65000
```

This setting takes effect after the OS is restarted.

After changing /etc/sysctl.conf, you can reflect the change instantly by executing the command below.

```
# sysctl -p
```

6.3.3 Avoiding insufficient ports

If a lot of servers and resources are used for EXPRESSCLUSTER, the number of temporary ports used for internal communications by EXPRESSCLUSTER may be insufficient and the servers may not work properly as the cluster server.
Adjust the range of port number and the time before a temporary port is released as needed.

6.3.4 Clock synchronization

In a cluster system, it is recommended to synchronize multiple server clocks regularly. Synchronize server clocks by using ntp.

6.3.5 NIC device name

Because of the ifconfig command specification, when the NIC device name is shortened, the length of the NIC device name which EXPRESSCLUSTER can handle depends on it.

6.3.6 Shared disk

- When you continue using the data on the shared disk at times such as server reinstallation, do not allocate a partition or create a file system.
- The data on the shared disk gets deleted if you allocate a partition or create a file system.
- EXPRESSCLUSTER controls the file systems on the shared disk. Do not include the file systems on the shared disk to /etc/fstab in operating system.
  (If the entry to is required /etc/fstab, please use the noauto option is not used ignore option.)
- Provide a disk heartbeat partition with 10 MB (10*1024*1024 bytes) or more of space. The disk heartbeat partition does not require any file system to be created.
- See the "Installation and Configuration Guide" for steps for shared disk configuration.

6.3.7 Mirror disk

- Set a management partition for mirror disk resource (cluster partition) and a partition for mirror disk resource (data partition).
- EXPRESSCLUSTER controls the file systems on mirror disks. Do not set the file systems on the mirror disks to /etc/fstab in operating system.
  (Do not enter a mirror partition device, mirror mount point, cluster partition, or data partition in /etc/fstab of the operating system.)
  (Do not enter /etc/fstab even with the ignore option specified.
  If you enter /etc/fstab with the ignore option specified, the entry will be ignored when mount is executed, but an error may subsequently occur when fsck is executed.)
  (Entering /etc/fstab with the noauto option specified is not recommended, either, because it may lead to an inadvertent manual mount or result in some application being mounted.)
- Provide the cluster partition with 1024 MiB or more of space. (Do not mind that specifying just 1024 MB actually provides more than the size due to a difference in the disk geometry.) Do not create any file system in the cluster partition.
- See the "Installation and Configuration Guide" for steps for mirror disk configuration.

6.3.8 Hybrid disk

- Configure the management partition (cluster partition) for hybrid disk resource and the partition used for hybrid disk resource (data partition).
- When a hybrid disk is allocated in the shared disk device, allocate the partition for the disk heart beat resource between servers sharing the shared disk device.
- EXPRESSCLUSTER controls the file systems on the hybrid disk. Do not include the file systems on the hybrid disk to /etc/fstab in operating system.

  (Do not enter a mirror partition device, mirror mount point, cluster partition, or data partition in /etc/fstab of the operating system.)

  (Do not enter /etc/fstab even with the ignore option specified.

  If you enter /etc/fstab with the ignore option specified, the entry will be ignored when mount is executed, but an error may subsequently occur when fsck is executed.)

  (Entering /etc/fstab with the noauto option specified is not recommended, either, because it may lead to an inadvertent manual mount or result in some application being mounted.)
- Provide the cluster partition with 1024 MiB or more of space. (Do not mind that specifying just 1024 MB actually provides more than the size due to a difference in the disk geometry.) Do not create any file system in the cluster partition.
- See the "Installation and Configuration Guide" for steps for hybrid disk configuration.
- When using this EXPRESSCLUSTER version, a file system must be manually created in a data partition used by a hybrid disk resource. For details about what to do when a file system is not created in advance, see "Settings after configuring hardware" in " Determining a system configuration" of the "Installation and Configuration Guide".

6.3.9 When using an ext3/ext4 file system for a mirror/hybrid disk resource

Block sizes

When creating an ext3/ext4 file system for the data partition of a mirror/hybrid disk resource by manually executing the mkfs command, avoid setting the block size at 1024.

The value, 1024, of the block size is not supported by mirror or hybrid disk resources. To explicitly use the block size, specify 2048 or 4096.

6.3.10 Adjusting OS startup time

It is necessary to configure the time from power-on of each node in the cluster to the server operating system startup to be longer than the following:

- The time from power-on of the shared disks to the point they become available.
- Heartbeat timeout time

  See the "Installation and Configuration Guide" for configuration steps.

6.3.11 Verifying the network settings

- The network used by Interconnect or Mirror disk connect is checked. It checks by all the servers in a cluster.
- See the "Installation and Configuration Guide" for configuration steps.

6.3.12 OpenIPMI

- The following functions use OpenIPMI.
- Final Action at Activation Failure / Deactivation Failure
- Monitor resource action upon failure
- User-mode monitor
- Shutdown monitor
- Forcibly stopping a physical machine
- OpenIPMI do not come with EXPRESSCLUSTER. You need to download and install the rpm packages for OpenIPMI.
- Check whether or not your server (hardware) supports OpenIPMI in advance.
- Note that even if the machine complies with ipmi standard as hardware, OpenIPMI may not run if you actually try to run them.
- If you are using a software program for server monitoring provided by a server vendor, do not choose ipmi as a monitoring method for user-mode monitor resource and shutdown stall monitor. Because these software programs for server monitoring and OpenIPMI both use BMC (Baseboard Management Controller) on the server, a conflict occurs preventing successful monitoring.

6.3.13 User mode monitor resource, shutdown monitoring (monitoring method: softdog)

- When softdog is selected as a monitoring method, use the soft dog driver.
  Make sure not to start the features that use the softdog driver except EXPRESSCLUSTER.
  Examples of such features are as follows:
  - Heartbeat feature that comes with OS
  - i8xx_tco driver
  - iTCO_WDT driver
  - watchdog feature and shutdown monitoring feature of systemd
- When softdog is selected as a monitoring method, make sure to set heartbeat that comes with OS not to start.
- When it sets softdog in a monitor method in SUSE LINUX 11, it is impossible to use with an i8xx_tco driver. When an i8xx_tco driver is unnecessary, make it the setting that i8xx_tco is not loaded.
- For Red Hat Enterprise Linux 6, when softdog is selected as a monitoring method, softdog cannot be used together with the iTCO_WDT driver. If the iTCO_WDT driver is not used, specify not to load iTCO_WDT.

6.3.14 Log collection

- The designated function of the generation of the syslog does not work by a log collection function in SUSE LINUX. The reason is because the suffixes of the syslog are different.
  Please change setting of rotate of the syslog as follows to use the appointment of the generation of the syslog of the log collection function.
- Please comment out "compress" and "date ext" of the /etc/logrotate.d/syslog file.
- When the total log size exceeds 2GB on each server, log collection may fail.

6.3.15 nsupdate and nslookup

- The following functions use nsupdate and nslookup.
  - Dynamic DNS resource of group resource (ddns)
  - Dynamic DNS monitor resource of monitor resource (ddnsw)
- EXPRESSCLUSTER does not include nsupdate and nslookup. Therefore, install the rmp files of nsupdate and nslookup, in addition to the EXPRESSCLUSTER installation.
- NEC does not support the items below regarding nsupdate and nslookup. Use nsupdate and nslookup at your own risk.
  - Inquiries about nsupdate and nslookup
  - Guaranteed operations of nsupdate and nslookup
  - Malfunction of nsupdate or nslookup or failure caused by such a malfunction
  - Inquiries about support of nsupdate and nslookup on each server

6.3.16 FTP monitor resources

- If a banner message to be registered to the FTP server or a message to be displayed at connection is long or consists of multiple lines, a monitor error may occur. When monitoring by the FTP monitor resource, do not register a banner message or connection message.

6.3.17 Notes on using Red Hat Enterprise Linux 7

- In mail reporting function takes advantage of the [mail] command of OS provides. Because the minimum composition is [mail] command is not installed, please execute one of the following.
- Select the [**SMTP**] by the **Mail Method** on the **Alert Service** tab of **Cluster Properties**.
- Installing mailx.

6.3.18 Notes on using Ubuntu

- To execute EXPRESSCLUSTER-related commands, execute them as the root user.
- Only a WebSphere monitor resource is supported in Application Server Agent. This is because other Application Server isn't supporting Ubuntu.
- In mail reporting function takes advantage of the [mail] command of OS provides. Because the minimum composition is [mail] command is not installed, please execute one of the following.
  - Select the [**SMTP**] by the **Mail Method** on the **Alert Service** tab of **Cluster Properties**.
  - Installing mailutils.
- Information acquisition by SNMP cannot be used.

6.3.19 Time synchronization in the AWS environtment

The following AWS-related resources execute the AWS CLI during activation, deactivation, or monitoring:

AWS Elastic IP resource
AWS Virtual IP resource
AWS Secondary IP resource
AWS DNS resource
AWS Elastic IP monitor resource
AWS Virtual IP monitor resource
AWS Secondary IP monitor resource
AWS AZ monitor resource
AWS DNS monitor resource
AWS Forced stop resource

If the date and time of an instance is not correctly set, executing the AWS CLI may fail due to the specification of AWS.
In such a case, correct the date and time of the instance by using a server such as an NTP server. For details, refer to "Setting the Time for Your Linux Instance" (https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/set-time.html)

6.3.20 IAM settings in the AWS environtment

This section describes the settings of IAM (Identity & Access Management) in AWS environment.
Some of EXPRESSCLUSTER's functions internally run AWS CLI for their processes. To run AWS CLI successfully, you need to set up IAM in advance.
You can give access permissions to AWS CLI by using IAM role or IAM user. IAM role method offers a high-level of security because you do not have to store AWS access key ID and AWS secret access key in an instance. Therefore, it is recommended to use IAM role basically.

The procedure of setting IAM is shown below.

1. First, create IAM policy by referring to "Creating IAM policy" explained below.
2. Next, set up the instance.
   To use IAM role, refer to "Setting up an instance by using IAM role" described later.
   To use IAM user, refer to "Setting up an instance by using IAM user" described later.

Creating IAM policy

Create a policy that describes access permissions for the actions to the services such as EC2 and S3 of AWS. The actions required for AWS-related resources and monitor resources to execute AWS CLI are as follows:

The necessary policies are subject to change.

- AWS virtual ip resource / AWS virtual ip monitor resource

Action	Description
ec2:DescribeNetworkInterfaces ec2:DescribeVpcs ec2:DescribeRouteTables	This is required for obtaining information of VPC, route table and network interfaces.
ec2:ReplaceRoute	This is required for updating the route table.

- AWS elastic ip resource /AWS elastic ip monitor resource

Action	Description
ec2:DescribeNetworkInterfaces ec2:DescribeAddresses	This is required for obtaining information of EIP and network interfaces.
ec2:AssociateAddress	This is required for associating EIP with ENI.
ec2:DisassociateAddress	This is required for disassociating EIP from ENI.

- AWS secondary ip resource / AWS secondary ip monitor resource

Action	Description
ec2:DescribeNetworkInterfaces ec2:DescribeSubnets	This is required for obtaining information on network interfaces and subnets.
ec2:AssignPrivateIpAddresses	This is required for assigning secondary IP addresses.
ec2:UnassignPrivateIpAddresses	This is required for deassigning secondary IP addresses.

- AWS AZ monitor resource

Action	Description
ec2:DescribeAvailabilityZones	This is required for obtaining information of the availability zone.

- AWS DNS resource / AWS DNS monitor resource

Action	Description
route53:ChangeResourceRecordSets	This is required for a resource record set is added or deleted or when the resource record set configuration is updated.
route53:GetChange	This is required for a resource record set is added or when the resource record set configuration is updated.
route53:ListResourceRecordSets	This is required for obtaining information of a resource record set.

- AWS forced stop resource

Action	Description
ec2:DescribeInstances	This is required for obtaining information on instances.
ec2:StopInstances	This is required for stopping instances.
ec2:RebootInstances	This is required for restarting instances.

- Function for sending data on the monitoring process time taken by the monitor resource, to Amazon CloudWatch.

Action	Description
cloudwatch:PutMetricData	This is required for sending custom metrics.

- Function for sending alert service messages to Amazon SNS

Action	Description
sns:Publish	This is required for sending messages.

The example of a custom policy as shown below permits actions used by all the AWS-related resources and monitor resources.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "ec2:Describe*",
                "ec2:ReplaceRoute",
                "ec2:AssociateAddress",
                "ec2:DisassociateAddress",
                "ec2:AssignPrivateIpAddresses",
                "ec2:UnassignPrivateIpAddresses",
                "ec2:StopInstances",
                "ec2:RebootInstances",
                "route53:ChangeResourceRecordSets",
                "route53:GetChange",
                "route53:ListResourceRecordSets"
            ],
            "Effect": "Allow",
            "Resource": "*"
        }
    ]
```

(continues on next page)

(continued from previous page)

```
}
```

You can create a custom policy from [Policies] - [Create Policy] in IAM Management Console

Setting up an instance by using IAM role

In this method, you can execute execute AWS CLI after creating IAM role and associate it with an instance.

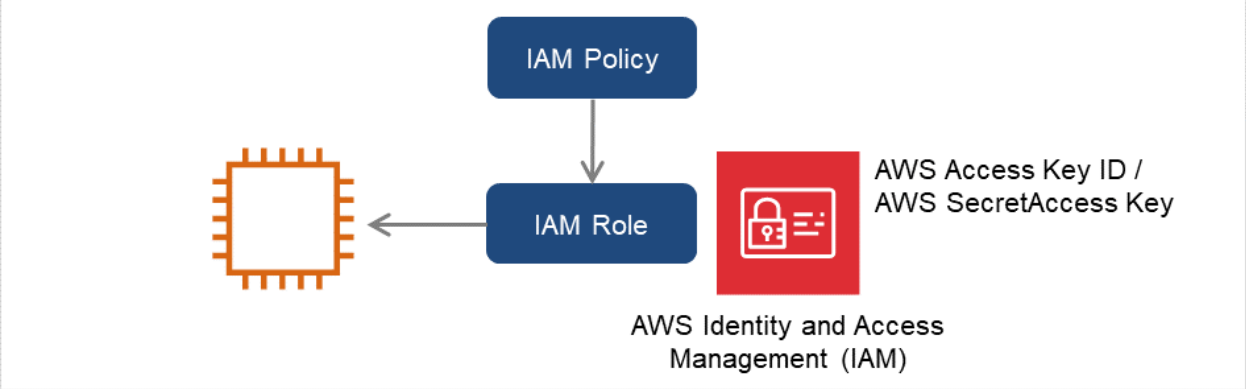


Fig. 6.7: Setting an instance by using IAM role

1) Create the IAM role and attach the IAM Policy to the role.

   You can create the IAM role from [Roles] - [Create New Role] in IAM Management Console

2) When creating an instance, specify the IAM role you created to **IAM Role**.

3) Log on to the instance.

4) Install Python.
   Install Python required by EXPRESSCLUSTER. First, confirm that Python has been installed on the machine. If not, install it by using the command such as the yum command. The installation path of the python command must be one of the following Use the python command initially found in the environment variable PATH.

   /sbin, /bin, /usr/sbin, /usr/bin

   If only Python 3 is installed and /usr/bin/python does not exist, create the symbolic link of /usr/bin/python for /usr/bin/python3.x (x indicates a version) or /usr/bin/python3.

5) Install the AWS CLI.

   The installation path of the AWS CLI must be any of the following:
   /sbin, /bin, /usr/sbin, /usr/bin, /usr/local/bin

   For details about how to set up the AWS CLI, refer to the following:
   http://docs.aws.amazon.com/cli/latest/userguide/installing.html

(If EXPRESSCLUSTER has been installed before installing Python or the AWS CLI, be sure to restart the OS before using EXPRESSCLUSTER.)

6) Execute the command from the shell as shown below

```
$ sudo aws configure
```

Input the information required to execute AWS CLI in response to the prompt. Do not input AWS access key ID and AWS secret access key.

```
AWS Access Key ID [None]: (Just press Enter key)
AWS Secret Access Key [None]: (Just press Enter key)
Default region name [None]: <default region name>
Default output format [None]: text
```

For "Default output format", other format than "text" may be specified.

If you input wrong information, delete the entire /root/.aws directory and execute the step described above.

Setting up an instance by using IAM user

In this method, you can execute execute AWS CLI after creating the IAM user and storing its access key ID and secret access key in the instance. You do not have to assign the IAM role to the instance when creating the instance.

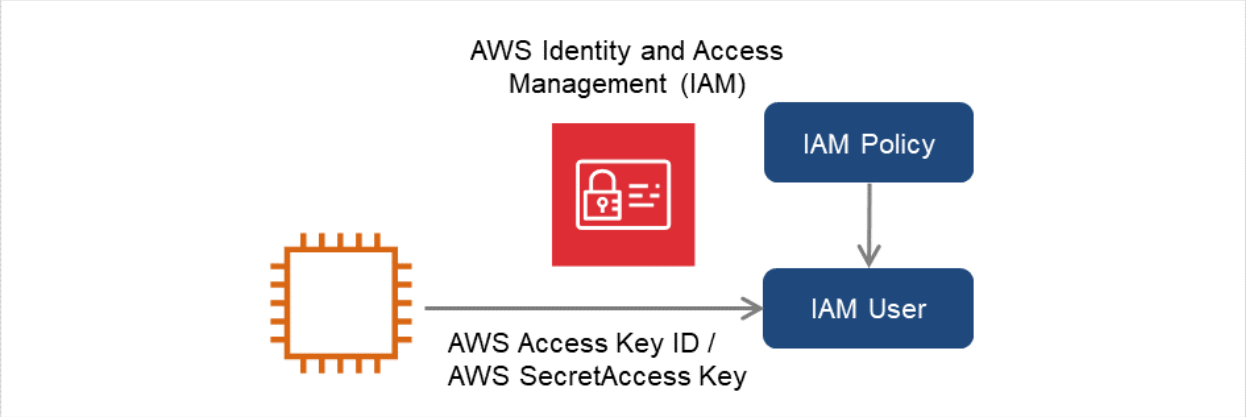


Fig. 6.8: Setting an instance by using IAM user

1) Create the IAM user and attach the IAM Policy to the role.
   You can create the IAM user in [Users] - [Create New Users] of IAM Management Console
2) Log on to the instance.
3) Install Python.
   Install Python required by EXPRESSCLUSTER. First, confirm that Python has been installed on the machine. If not, install it by using the command such as the yum command.
   The installation path of the python command must be one of the following Use the python command initially found in the environment variable PATH.

   /sbin, /bin, /usr/sbin, /usr/bin

If only Python 3 is installed and /usr/bin/python does not exist, create the symbolic link of /usr/bin/python for /usr/bin/python3.*x* (*x* indicates a version) or /usr/bin/python3.

4) Install the AWS CLI.

The installation path of the AWS CLI must be any of the following:
/sbin, /bin, /usr/sbin, /usr/bin, /usr/local/bin

For details about how to set up the AWS CLI, refer to the following:
http://docs.aws.amazon.com/cli/latest/userguide/installing.html

(If EXPRESSCLUSTER has been installed before installing Python or the AWS CLI, be sure to restart the OS before using EXPRESSCLUSTER.)

5) Execute the command from the shell as shown below

```
$ sudo aws configure
```

Input the information required to execute AWS CLI in response to the prompt. Obtain AWS access key ID and AWS secret access key from IAM user detail screen to input.

```
AWS Access Key ID [None]: <AWS access key>
AWS Secret Access Key [None]: <AWS secret access key>
Default region name [None]: <default region name >
Default output format [None]: text
```

For "Default output format", other format than "text" may be specified.
If you input wrong information, delete the entire /root/.aws directory and execute the step described above.

6.3.21 Azure DNS resources

- For the procedures to install Azure CLI and create a service principal, refer to the "EXPRESSCLUSTER X HA Cluster Configuration Guide for Microsoft Azure (Linux)".
- The Azure CLI and Python must be installed because the Azure DNS resource uses them. Python is supplied with an OS such as Red Hat Enterprise Linux and Cent OS. For details about the Azure CLI, refer to the following website:

Microsoft Azure document:
https://docs.microsoft.com/en-us/azure/

- The Azure DNS service must be installed because the Azure DNS resource uses it. For details about Azure DNS, refer to the following website:

Azure DNS:

https://azure.microsoft.com/en-us/services/dns/

- To set up EXPRESSCLUSTER to work with Microsoft Azure, a Microsoft Azure organizational account is required. An account other than the organizational account cannot be used because an interactive login is required when executing the Azure CLI.
- It is necessary to create a service principal with Azure CLI.

The Azure DNS resource logs in to Microsoft Azure and performs the DNS zone registration. The Azure DNS resource uses Azure login based on service principal when logging in Microsoft Azure.
For details about a service principal and procedure, refer to the following websites:

Log in with Azure CLI 2.0:
https://docs.microsoft.com/en-us/cli/azure/authenticate-azure-cli?view=azure-cli-latest

Create an Azure service principal with Azure CLI 2.0:
https://docs.microsoft.com/en-us/cli/azure/create-an-azure-service-principal-azure-cli?view=azure-cli-latest

When changing the role of the created service principal from the default role "Contributor" to another role, select the role that can access all of the following operations as the Actions properties.
If the role is changed to one that does not meet this condition, starting the Azure DNS resource fails due to an error.

For Azure CLI 1.0:

Microsoft.Network/dnsZones/read
Microsoft.Network/dnsZones/A/write
Microsoft.Network/dnsZones/A/read
Microsoft.Network/dnsZones/A/delete
Microsoft.Network/dnsZones/NS/read

For Azure CLI 2.0:

Microsoft.Network/dnsZones/A/write
Microsoft.Network/dnsZones/A/delete
Microsoft.Network/dnsZones/NS/read

- Azure Private DNS is not supported.

6.3.22 Google Cloud DNS resources

- Google Cloud DNS resources use Cloud DNS by Google Cloud. For the details on Cloud DNS, refer to the following website.

  Cloud DNS
  https://cloud.google.com/dns/

- Cloud SDK needs to be installed to operate Cloud DNS. For the details on Cloud SDK, refer to the following website.

  Cloud SDK
  https://cloud.google.com/sdk/

- Cloud SDK needs to be authorized by the account with the permissions for the API methods below:

  dns.changes.create
  dns.changes.get
  dns.managedZones.get
  dns.resourceRecordSets.create
  dns.resourceRecordSets.delete
  dns.resourceRecordSets.list
  dns.resourceRecordSets.update

  As for authorizing Cloud SDK, refer to the following website.

  Authorizing Cloud SDK tools
  https://cloud.google.com/sdk/docs/authorizing

6.3.23 Samba monitor resources

- In order to support SMB protocol version 2.0 or later, NTLM authentication, and SMB signature, Samba monitor resources use a shared library 'libsmbclient.so.0' for the internal version 4.1.0-1 or later. Confirm that it is installed since libsmbclient.so.0 is included in libsmbclient package.
- If the version of libsmbclient is 3 or earlier (for example, libsmbclient included in RHEL 6), .you can specify only either 139 or 445 for **Port**. Specify the port number included in smb ports of smb.conf.
- The version of SMB protocol supported by Samba monitor resource depends on the installed libsmbclient. You can confirm whether to receive supports from libsmbclient by testing a connection to shared area of the monitoring target by using the smbclient command which each distributer provides.

6.3.24 About HTTP network partition resolution resources and Witness heartbeat resources

- For HTTP network partition resolution resources and Witness heartbeat resources, using SSL requires OpenSSL 1.0/1.1. By default, the following libraries are used:
  - libssl.so.10 (if you installed the rpm package of EXPRESSCLUSTER)
  - libssl.so.1.0.0 (if you installed the deb package of EXPRESSCLUSTER)

  To use other libraries, go to the **Encryption** tab and set **SSL Library** and **Crypto Library**.

6.3.25 CLI settings in the OCI environtment

This section describes the settings of CLI in OCI environment.
Some of EXPRESSCLUSTER's functions internally run OCI CLI for their processes.
To run OCI CLI successfully, you need to set up in advance.

For OCI CLI settings, refer to the following website.

Oracle Cloud Infrastructure Documentation - Command Line Interface (CLI)
https://docs.oracle.com/en-us/iaas/Content/API/Concepts/cliconcepts.htm

6.3.26 Configuring OCI forced-stop resource

Using the OCI forced-stop resource requires changing parameter values in the following script, based on the directory where OCI CLI commands are installed and on the location where the OCI configuration file is stored:

```
<EXPRESSCLUSTER installation path>/cloud/oci/clpociforcestop.sh

-   Parameter value to be changed according to the directory where OCI CLI␣
↪commands are installed
    export PATH=$PATH:[directory where OCI CLI commands are installed]
    Example: export PATH=$PATH:/root/bin

-   Parameter value to be changed according to the location where the OCI␣
↪configuration file is stored
    OCI_Path="[path to the OCI configuration file]"
    Example: OCI_Path="/root/.oci/config"
```

6.4 Notes when creating EXPRESSCLUSTER configuration data

Notes when creating a cluster configuration data and before configuring a cluster system is described in this section.

6.4.1 Directories and files in the location pointed to by the EXPRESSCLUSTER installation path

The directories and files in the location pointed to by the EXPRESSCLUSTER installation path must not be handled (edited, created, added, or deleted) by using any application or tool other than EXPRESSCLUSTER.
Any effect on the operation of a directory or file caused by using an application or tool other than EXPRESSCLUSTER will be outside the scope of NEC technical support.

6.4.2 Environment variable

The following processes cannot be executed in an environment in which more than 255 environment variables are set. When using the following function of resource, set the number of environmental variables less than 256.

- Group start/stop process
- Start/Stop script executed by EXEC resource when activating/deactivating
- Script executed by Custom monitor Resource when monitoring
- Script before final action after the group resource or the monitor resource error is detected
- Script to be executed before and after activating or deactivating a group resource
- The script for forced stop

Note: The total number of environment variables set in the system and EXPRESSCLUSTER must be less than 256. About 30 environment variables are set in EXPRESSCLUSTER.

6.4.3 Server reset, server panic and power off

When EXPRESSCLUSTER performs "Server Reset", "Server Panic," or "Server power off", servers are not shut down normally. Therefore, the following may occur.

- Damage to a mounted file system
- Loss of unsaved data
- Suspension of OS dump collection

"Server reset" or "Server panic" occurs in the following settings:

- Action at an error occurred when activating/inactivating group resources
  - Sysrq Panic
  - Keepalive Reset
  - Keepalive Panic
  - BMC Reset
  - BMC Power Off

  - BMC Power Cycle
  - BMC NMI
- Final action at detection of an error in monitor resource
  - Sysrq Panic
  - Keepalive Reset
  - Keepalive Panic
  - BMC Reset
  - BMC Power Off
  - BMC Power Cycle
  - BMC NMI
- Action at detection of user mode monitor timeout
  - Monitoring method softdog
  - Monitoring method ipmi
  - Monitoring method keepalive

Note: "Server panic" can be set only when the monitoring method is "keepalive."

- Shutdown stall mentoring
  - Monitoring method softdog
  - Monitoring method ipmi
  - Monitoring method keepalive

Note: "Server panic" can be set only when the monitoring method is "keepalive."

- Operation of Forced Stop
  - BMC reset
  - BMC power off
  - BMC cycle
  - BMC NMI
  - VMware vSphere power off

6.4.4 Final action for group resource deactivation error

If you select **No Operation** as the final action when a deactivation error is detected, the group does not stop but remains in the deactivation error status. Make sure not to set **No Operation** in the production environment.

6.4.5 Verifying raw device for VxVM

Check the raw device of the volume raw device in advance:

1. Import all disk groups which can be activated on one server and activate all volumes before installing EXPRESSCLUSTER.
2. Run the command below:
   In the following output example, `/dev/raw/raw2` and `/dev/raw/raw3` are the raw device names.
   In addition, the `major ..., minor ...` part represents major/minor numbers.

   ```
   # raw -qa
   /dev/raw/raw2: bound to major 199, minor 2
   /dev/raw/raw3: bound to major 199, minor 3
   ```

   Example: Assuming the disk group name and volume name are:

   - Disk group name: dg1
   - Volume name under dg1: vol1, vol2

3. Run the command below:
   In the following output example, the `199, 2` and `199, 3` parts represent major/minor numbers.

   ```
   # ls -l /dev/vx/dsk/dg1/
   brw------- 1 root root 199, 2 May 15 22:13 vol1
   brw------- 1 root root 199, 3 May 15 22:13 vol2
   ```

4. Confirm that major and minor numbers are identical between step 2 and step 3.

   Make sure not to set the raw devices confirmed in Step 2 to the disk heartbeat resource, the disk resource (disk type is other than "VxVM"), or the disk monitor resource (monitoring method is other than READ (VxVM)) of EXPRESSCLUSTER.

6.4.6 Selecting Mirror disk file system

Following is the currently supported file systems:

- ext3
- ext4
- xfs
- reiserfs
- jfs
- vxfs
- none(no file system)

6.4.7 Selecting Hybrid disk file system

The following are the currently supported file systems:

- ext3
- ext4
- xfs
- reiserfs
- none(no file system)

6.4.8 Time to start a single serve when many Mirror disks are defined.

If many mirror disk resources are defined and a short time is set to **Time to wait for the other servers to start up**, it may take time to start a mirror agent and mirror disk resources and monitor resources related to mirror disks may not start properly when a single server is started.
If such an event occurs when starting a single server, change the value set to the time to wait for synchronization to a large value (by selecting **Cluster Properties** - **Timeout** tab - **Server Sync Wait Time**).

6.4.9 RAW monitoring of Disk monitor resources

- When raw monitoring of disk monitor resources is set up, partitions cannot be monitored if they have been or will possibly be mounted. These partitions cannot be monitored even if you set device name to "whole device" (device indicating the entire disks).
- Allocate a partition dedicated to monitoring and set up the partition to use the raw monitoring of disk monitor resources.

6.4.10 Delay warning rate

If the delay warning rate is set to 0 or 100, the following can be achieved:

- When 0 is set to the delay monitoring rate
  An alert for the delay warning is issued at every monitoring.
  By using this feature, you can calculate the polling time for the monitor resource at the time the server is heavily loaded, which will allow you to determine the time for monitoring time-out of a monitor resource.
- When 100 is set to the delay monitoring rate
  The delay warning will not be issued.
  Be sure not to set a low value, such as 0%, except for a test operation.

6.4.11 Disk monitor resource (monitoring method TUR)

- You cannot use the TUR methods on a disk or disk interface (HBA) that does not support the Test Unit Ready (TUR) and SG_IO commands of SCSI. Even if your hardware supports these commands, consult the driver specifications because the driver may not support them.
- S-ATA disk interface may be recognized as IDE disk interface (hd) or SCSI disk interface (sd) by OS depending on disk controller type and distribution. When it is recognized as IDE interface, all TUR methods cannot be used. If it is recognized as SCSI disk interface, TUR (legacy) can be used. Note that TUR (generic) cannot be used.
- TUR methods burdens OS and disk load less compared to Read methods.
- In some cases, TUR methods may not be able to detect errors in I/O to the actual media.

6.4.12 Configuring LAN heartbeat resources or kernel mode LAN heartbeat resources

- For an interconnect with the highest priority, configure LAN heartbeat resources or kernel mode LAN heartbeat resources which can be exchanged between all servers.
- Configuring at least two kernel mode LAN heartbeat resources is recommended unless it is difficult to add a network to an environment such as the cloud or a remote cluster.
- It is recommended to register both an interconnect-dedicated LAN and a public LAN as LAN heartbeat resources.
- It is recommended to use kernel mode LAN heartbeat resource for distribution kernel of which kernel mode LAN heartbeat can be used.

6.4.13 Double-byte character set that can be used in script comments

- Scripts edited in Linux environment are dealt as EUC code, and scripts edited in Windows environment are dealt as Shift-JIS code. In case that other character codes are used, character corruption may occur depending on environment.

6.4.14 The character code and line feed code in a script

- If you use the clpcfctrl command to apply the settings of a script created by some means other than Cluster WebUI, make sure beforehand that the character code and line feed code in the script are the same as those in the configuration data file (clp.conf). If the character code or the line feed code is different between the script and clp.conf, the script may not work properly.

6.4.15 System monitor resource settings

- Pattern of detection by resource monitoring

  The System Resource Agent detects by using thresholds and monitoring duration time as parameters.

  The System Resource Agent collects the data (number of opened files, number of user processes, number of threads, used size of memory, CPU usage rate, and used size of virtual memory) on individual system resources continuously, and detects errors when data keeps exceeding a threshold for a certain time (specified as the duration time).

6.4.16 Message receive monitor resource settings

- Error notification to message receive monitor resources can be done in any of three ways: using the clprexec command, or linkage with the server management infrastructure.
- To use the clprexec command, use the relevant file stored on the EXPRESSCLUSTER CD. Use this method according to the OS and architecture of the notification-source server. The notification-source server must be able to communicate with the notification-destination server.
- For the linkage with the server management infrastructure, see "Linkage with Server Management Infrastructure" in the "Hardware Feature Guide".

6.4.17 JVM monitor resource settings

- When the monitoring target is the WebLogic Server, the maximum values of the following JVM monitor resource settings may be limited due to the system environment (including the amount of installed memory):
  - **The number** under **Monitor the requests in Work Manager**
  - **Average** under **Monitor the requests in Work Manager**
  - **The number** of **Waiting Requests** under **Monitor the requests in Thread Pool**
  - **Average** of **Waiting Requests** under **Monitor the requests in Thread Pool**
  - **The number** of **Executing Requests** under **Monitor the requests in Thread Pool**
  - **Average** of **Executing Requests** under **Monitor the requests in Thread Pool**
- When the monitoring-target is a 64-bit JRockit JVM, the following parameters cannot be monitored because the maximum amount of memory acquired from the JRockit JVM is a negative value that disables the calculation of the memory usage rate:
  - **Total Usage** under **Monitor Heap Memory Rate**
  - **Nursery Space** under **Monitor Heap Memory Rate**
  - **Old Space** under **Monitor Heap Memory Rate**
  - **Total Usage** under **Monitor Non-Heap Memory Rate**
  - **Class Memory** under **Monitor Non-Heap Memory Rate**
- To use the JVM monitor resources, install the Java runtime environment (JRE) described in "**Operation environment for JVM monitor**" in "4. *Installation requirements for EXPRESSCLUSTER*" You can use either the same JRE as that used by the monitoring target (WebLogic Server or WebOTX) or a different JRE.
- The monitor resource name must not include a blank.

6.4.18 EXPRESSCLUSTER startup when using volume manager resources

- When EXPRESSCLUSTER starts up, the system startup may take some time because of the deactivation processing performed by the vgchange command if the volume manager is lvm or the deport processing if it is vxvm. If this presents a problem, edit the startup or stop script of the EXPRESSCLUSTER main body as shown below.
  - For an init.d environment, edit /etc/init.d/clusterpro as shown below.

```
#!/bin/sh
#
# Startup script for the EXPRESSCLUSTER daemon
#
                            :
                            :
# See how we were called.
case "$1" in
  start)
                            :
                            :
        # export all volmgr resource
#         clp_logwrite "$1" "clpvolmgrc start." init_main
#         ./clpvolmgrc -d > /dev/null 2>&1
#         retvolmgrc=$?
#         clp_logwrite "$1" "clpvolmgrc end.("$retvolmgrc")" init_main
                            :
                            :
```

– For a systemd environment, edit /opt/nec/clusterpro/etc/systemd/clusterpro.sh as shown below.

```
#!/bin/sh
#
# Startup script for the EXPRESSCLUSTER daemon
#
                            :
                            :
# See how we were called.
case "$1" in
  start)
                            :
                            :
        # export all volmgr resource
#         clp_logwrite "$1" "clpvolmgrc start." init_main
#         ./clpvolmgrc -d > /dev/null 2>&1
#         retvolmgrc=$?
#         clp_logwrite "$1" "clpvolmgrc end.("$retvolmgrc")" init_main
```

6.4.19 Setting up AWS elastic ip resources

- IPv6 is not supported.
- In the AWS environment, floating IP resources, floating IP monitor resources, virtual IP resources, and virtual IP monitor resources cannot be used.
- Only ASCII characters is supported. Check that the character besides ASCII character isn't included in an execution result of the following command.

  `aws ec2 describe-addresses --allocation-ids <EIP ALLOCATION ID>`
- AWS elastic IP resources associate an EIP with the primary private IP address of an ENI, but not with its secondary private IP address.

6.4.20 Setting up AWS virtual ip resources

- IPv6 is not supported.
- In the AWS environment, floating IP resources, floating IP monitor resources, virtual IP resources, and virtual IP monitor resources cannot be used.
- Only ASCII characters is supported. Check that the character besides ASCII character isn't included in an execution result of the following command.

```
aws ec2 describe-vpcs --vpc-ids <VPC ID>
aws ec2 describe-route-tables --filters Name=vpc-id,Values=<VPC ID>
aws ec2 describe-network-interfaces --network-interface-ids <ENI ID>
```

- AWS virtual IP resources cannot be used if access via a VPC peering connection is necessary. This is because it is assumed that an IP address to be used as a VIP is out of the VPC range and such an IP address is considered invalid in a VPC peering connection. If access via a VPC peering connection is necessary, use the AWS DNS resource that use Amazon Route 53.
- An AWS virtual IP resource starts up normally, even if the route table to be used by instances does not include any route to an IP address to be used by the AWS virtual IP resource. This operation is as required. When activated, an AWS virtual IP resource updates the content of a route table that includes a specified IP address entry. Finding no route table, the resource considers the situation as nothing to be updated and therefore as normal. Which route table should have a specified entry, depending on the system configuration, is not the resource's criterion for judging the normality.

6.4.21 Setting up AWS secondary ip resources

- IPv6 is not supported.
- In the AWS environment, floating IP resources, floating IP monitor resources, virtual IP resources, and virtual IP monitor resources cannot be used.
- Only ASCII characters is supported. Check that the character besides ASCII character isn't included in an execution result of the following command.

```
aws ec2 describe-network-interfaces --network-interface-ids <ENI ID>
aws ec2 describe-subnets --subnet-ids <SUBNET_ID>
```

- No AWS secondary IP resources can be used in a configuration with a different subnet.
- The number of secondary IP addresses to be assigned for AWS secondary IP resources has an upper limit for each instance type.

  For more information, refer to the following:
  https://docs.aws.amazon.com/en_en/AWSEC2/latest/UserGuide/using-eni.html#AvailableIpPerENI

6.4.22 Setting up AWS DNS resources

- IPv6 is not supported.
- In the AWS environment, floating IP resources, floating IP monitor resources, virtual IP resources, and virtual IP monitor resources cannot be used.
- In the **Resource Record Set Name** field, enter a name without an escape code. **If it is included in the Resource Record Set Name**, a monitor error occurs.

- When activated, an AWS DNS resource does not await the completion of propagating changed DNS settings to all Amazon Route 53 DNS servers. This is due to the specification of Route 53: It takes time for the changes of a resource record set to be propagated throughout the network. Refer to "*Setting up AWS DNS monitor resources*".
- Associated with a single account, an AWS DNS resource cannot be used for different accounts, AWS access key IDs, or AWS secret access keys. If you want such usage, consider using a script (EXEC resource) to execute the AWS CLI.

6.4.23 Setting up AWS DNS monitor resources

- The AWS DNS monitor resource runs the AWS CLI for monitoring. The AWS DNS monitor resource uses **AWS CLI timeout** set to the AWS DNS resource as the timeout of the AWS CLI execution.
- Immediately after the AWS DNS resource is activated, monitoring by the AWS DNS monitor resource may fail due to the following events. If monitoring failed, set **Wait Time to Start Monitoring** of the AWS DNS monitor resource longer than the time to reflect the changed DNS setting of Amazon Route 53 (https://aws.amazon.com/route53/faqs/).
  1. When the AWS DNS resource is activated, a resource record set is added or updated.
  2. If the AWS DNS monitor resource starts monitoring before the changed DNS setting of Amazon Route 53 is applied, name resolution cannot be done and monitoring fails.
     The AWS DNS monitor resource will continue to fail monitoring while a DNS resolver cache is enabled.
  3. The changed DSN setting of Amazon Route 53 is applied.
  4. Name resolution succeeds after the **TTL** valid period of the AWS DNS resource elapses. Then, the AWS DNS monitor resource succeeds monitoring.

6.4.24 Setting up Azure probe port resources

- IPv6 is not supported.
- In the Microsoft Azure environment, floating IP resources, floating IP monitor resources, virtual IP resources, and virtual IP monitor resources cannot be used.

6.4.25 Setting up Azure load balance monitor resources

- When a Azure load balance monitor resource error is detected, there is a possibility that switching of the active server and the stand-by server from Azure load balancer is not performed correctly. Therefore, in the Final Action of Azure load balance monitor resources and the recommended that you select Stop the cluster service and shutdown OS.

6.4.26 Setting up Azure DNS resources

- IPv6 is not supported.
- In the Microsoft Azure environment, floating IP resources, floating IP monitor resources, virtual IP resources, and virtual IP monitor resources cannot be used.

6.4.27 Setting up Google Cloud virtual IP resources

- IPv6 is not supported.

6.4.28 Setting up Google Cloud load balance monitor resources

- For **Final Action** of Google Cloud load balance monitor resources, selecting **Stop cluster service and shutdown OS** is recommended. When a Google Cloud load balance monitor resource detects an error, the load balancer may not correctly switch between the active server and the standby server.

6.4.29 Setting up Google Cloud DNS resources

- IPv6 is not supported.
- In the Google Cloud Platform environment, floating IP resources, floating IP monitor resources, virtual IP resources, and virtual IP monitor resources cannot be used.
- When using multiple Google Cloud DNS resources in the cluster, you need to configure them to prevent their simultaneous activation/deactivation for their dependence or a wait for a group start/stop. Their simultaneous activation/deactivation may cause an error.

6.4.30 Setting up Oracle Cloud virtual IP resources

- IPv6 is not supported.

6.4.31 Setting up Oracle Cloud load balance monitor resources

- For **Final Action** of Oracle Cloud load balance monitor resources, selecting **Stop cluster service and shutdown OS** is recommended. When an Oracle Cloud load balance monitor resource detects an error, the load balancer may not correctly switch between the active server and the standby server.

6.4.32 Coexistence of a mirror disk resource with a hybrid disk resource

- A mirror disk resource and a hybrid disk resource cannot coexist in the same failover group.

6.5 After starting operating EXPRESSCLUSTER

Notes on situations you may encounter after start operating EXPRESSCLUSTER are described in this section.

6.5.1 Error message in the load of the mirror driver in an environment such as udev

In the load of the mirror driver in an environment such as udev, logs like the following may be recorded into the message file:

```
kernel: [I] <type: liscal><event: 141> NMP1 device does not exist. (liscal_
↪make_request)
kernel: [I] <type: liscal><event: 141> - This message can be recorded on udev␣
↪environment when liscal is initializing NMPx.
kernel: [I] <type: liscal><event: 141> - Ignore this and following messages␣
↪'Buffer I/O error on device NMPx' on udev environment.

kernel: Buffer I/O error on device NMP1, logical block 0

kernel: <liscal liscal_make_request> NMP1 device does not exist.
kernel: Buffer I/O error on device NMP1, logical block 112
```

This phenomenon is not abnormal.
When you want to prevent the output of the error message in the udev environment, add the following file in /etc/udev/rules.d.
Note, however, that error messages may be output even if the rule files are added in Red Hat Enterprise Linux 7 or Asianux Server 7.

filename: 50-liscal-udev.rules

```
ACTION=="add", DEVPATH=="/block/NMP*", OPTIONS+="ignore_device"
ACTION=="add", DEVPATH=="/devices/virtual/block/NMP*", OPTIONS+="ignore_device"
```

6.5.2 Buffer I/O error log for the mirror partition device

If the mirror partition device is accessed when a mirror disk resource or hybrid disk resource is inactive, log messages such as the ones shown below are recorded in the messages file.

```
kernel: [W] <type: liscal><event: 144> NMPx I/O port has been closed,␣
↪mount(0), io(0). (PID=xxxxx)
kernel: [I] <type: liscal><event: 144> - This message can be recorded on␣
↪hotplug service starting when NMPx is not active.
kernel: [I] <type: liscal><event: 144> - This message can be recorded by fsck␣
↪command when NMPx becomes active.
kernel: [I] <type: liscal><event: 144> - Ignore this and following messages␣
↪'Buffer I/O error on device NMPx' on such environment.

:

kernel: Buffer I/O error on device /dev/NMPx, logical block xxxx
kernel: [W] <type: liscal><event: 144> NMPx I/O port has been closed,␣
↪mount(0), io(0). (PID=xxxx)
```

:

```
kernel: [W] <type: liscal><event: 144> NMPx I/O port has been closed,␣
↪mount(0), io(0). (PID=xxxx)
```

```
kernel: <liscal liscal_make_request> NMPx I/O port is close, mount(0), io(0).
kernel: Buffer I/O error on device /dev/NMPx, logical block xxxx
```

(Where *x* and *xxxx* each represent a given number.)

The possible causes of this phenomenon are described below.
(In the case of a hybrid disk resource, the term "mirror disk resource" should be replaced with "hybrid disk resource" hereinafter.)

- When the udev environment is responsible
  - In this case, when the mirror driver is loaded, the message "kernel: Buffer I/O error on device /dev/NMP*x*, logical block *xxxx*" is recorded together with the message "kernel: [I] <type: liscal><event: 141>".
  - These messages do not indicate any error and have no impact on the operation of EXPRESSCLUSTER.
  - For details, see "Error message in the load of the mirror driver in an environment such as udev" in this chapter.
- When an information collection command (sosreport, sysreport, blkid, etc.) of the operating system has been executed
  - In this case, these messages do not indicate any error and have no impact on the operation of EXPRESSCLUSTER.
  - When an information collection command provided by the operating system is executed, the devices recognized by the operating system are accessed. When this occurs, the inactive mirror disk is also accessed, resulting in the above messages being recorded.
  - There is no way of suppressing these messages by using the settings of EXPRESSCLUSTER or other means.
- When the unmount of the mirror disk has timed out
  - In this case, these messages are recorded together with the message that indicates that the unmount of the mirror disk resource has timed out.
  - EXPRESSCLUSTER performs the "recovery operation for the detected deactivation error" of the mirror disk resource. It is also possible that there is inconsistency in the file system.
  - For details, see "6.5.3. *Cache swell by a massive I/O*" in this chapter.
- When the mirror partition device may be left mounted while the mirror disk is inactive
  - In this case, the above messages are recorded after the following actions are taken.
    (1) After the mirror disk resource is activated, the user or an application (for example, NFS) specifies an additional mount in the mirror partition device (/dev/NMP*x*) or the mount point of the mirror disk resource.
    (2) Then, the mirror disk resource is deactivated without unmounting the mount point added in (1).
  - While the operation of EXPRESSCLUSTER is not affected, it is possible that there is inconsistency in the file system.

  - For details, see "6.5.4. *When multiple mounts are specified for a resource like a mirror disk resource*" in this chapter.

- When multiple mirror disk resources are configured
  - With some distributions, when two or more mirror disk resources are configured, the above messages may be output due to the behavior of fsck if the resources are active.
  - For details, see "6.5.5. *Messages written to syslog when multiple mirror disk resources or hybrid disk resources are used*"
- When the mirror disk resource is accessed by a certain application
  - Besides the above cases, it is possible that a certain application has attempted to access the inactive mirror disk resource.
  - When the mirror disk resource is not active, the operation of EXPRESSCLUSTER is not affected.

6.5.3 Cache swell by a massive I/O

- In case that a massive amount of write over the disk capability to the mirror disk resource or the hybrid disk resource are executed, even though the mirror connection is alive, the control from write may not return or memory allocation failure may occur.

  In case that a massive amount of I/O requests over transaction performance exist, and then the file system ensure a massive amount of cache and the cache or the memory for the user space (HIGHMEM zone) are insufficient, the memory for the kernel space (NORMAL zone) may be used.

  Change the settings so that the parameter will be changed at OS startup by using sysctl or other commands.

```
/proc/sys/vm/lowmem_reserve_ratio
```

- In case that a massive amount of accesses to the mirror disk resource or the hybrid disk resource are executed, it may take much time before the cache of the file systems is written out to the disks when unmounting at disk resource deactivation.

  If, at this moment, the unmounting times out before the writing from the file system to the disks is completed, I/O error messages or unmount failure messages like those shown below may be recorded.

  In this case, change the unmount timeout length for the disk resource in question to an adequate value such that the writing to the disk will be normally completed.

  Example 1:

```
expresscls: [I] <type: rc><event: 40> Stopping mdx resource has␣
↪started.
kernel: [I] <type: liscal><event: 193> NMPx close I/O port OK.
kernel: [I] <type: liscal><event: 195> NMPx close mount port OK.
kernel: [I] <type: liscal><event: 144> NMPx I/O port has been closed,␣
↪mount(0), io(0).
kernel: [I] <type: liscal><event: 144> - This message can be recorded␣
↪on hotplug service starting when NMPx is not active.
kernel: [I] <type: liscal><event: 144> - This message can be recorded␣
↪by fsck command when NMPx becomes active.
kernel: [I] <type: liscal><event: 144> - Ignore this and following␣
↪messages 'Buffer I/O error on device NMPx' on such environment.
kernel: Buffer I/O error on device NMPx, logical block xxxx
kernel: [I] <type: liscal><event: 144> NMPx I/O port has been closed,␣
↪mount(0), io(0).
kernel: Buffer I/O error on device NMPx, logical block xxxx
```

:

Example 2:

```
expresscls: [I] <type: rc><event: 40> Stopping mdx resource has␣
↪started.
kernel: [I] <type: liscal><event: 148> NMPx holder 1. (before umount)
expresscls: [E] <type: md><event: 46> umount timeout. Make sure that␣
↪the length of Unmount Timeout is appropriate. (Device:mdx)

:

expresscls: [E] <type: md><event: 4> Failed to deactivate mirror disk.
↪ Umount operation failed.(Device:mdx)
kernel: [I] <type: liscal><event: 148> NMPx holder 1. (after umount)
expresscls: [E] <type: rc><event: 42> Stopping mdx resource has␣
↪failed.(83 : System command timeout (umount, timeout=xxx))

:
```

6.5.4 When multiple mounts are specified for a resource like a mirror disk resource

- If, after activation of a mirror disk resource or hybrid disk resource, you have created an additional mount point in a different location by using the mount command for the mirror partition device (/dev/NMP*x*) or the mount point (or a part of the file hierarchy for the mount point), you must unmount that additional mount point before the disk resource is deactivated.

  If the deactivation is performed without the additional mount point being unmounted, the file system data remaining in memory may not be completely written out to the disks. As a result, the I/O to the disks is closed and the deactivation is completed although the data on the disks are incomplete.

  Because the file system will still try to continue writing to the disks even after the deactivation is completed, I/O error messages like those shown below may be recorded.

  After this, an attempt to stop the mirror agent, such as when stopping the server, will fail, since the mirror driver cannot be terminated. This may cause the server to restart.

  Example:

```
expresscls: [I] <type: rc><event: 40> Stopping mdx resource has␣
↪started.
kernel: [I] <type: liscal><event: 148> NMP1 holder 1. (before umount)
kernel: [I] <type: liscal><event: 148> NMP1 holder 1. (after umount)
kernel: [I] <type: liscal><event: 193> NMPx close I/O port OK.
kernel: [I] <type: liscal><event: 195> NMPx close mount port OK.
expresscls: [I] <type: rc><event: 41> Stopping mdx resource has␣
↪completed.
kernel: [I] <type: liscal><event: 144> NMPx I/O port has been closed,␣
↪mount(0), io(0).
kernel: [I] <type: liscal><event: 144> - This message can be recorded␣
↪on hotplug service starting when NMPx is not active.
kernel: [I] <type: liscal><event: 144> - This message can be recorded␣
↪by fsck command when NMPx becomes active.
kernel: [I] <type: liscal><event: 144> - Ignore this and following␣
↪messages 'Buffer I/O error on device NMPx' on such environment.
kernel: Buffer I/O error on device NMPx, logical block xxxxx
kernel: lost page write due to I/O error on NMPx
```

```
kernel: [I] <type: liscal><event: 144> NMPx I/O port has been closed,␣
↪mount(0), io(0).
kernel: Buffer I/O error on device NMPx, logical block xxxxx
kernel: lost page write due to I/O error on NMPx

:
```

6.5.5 Messages written to syslog when multiple mirror disk resources or hybrid disk resources are used

When more than two mirror disk resources or hybrid disk resources are configured on a cluster, the following messages may be written to the OS message files when the resources are activated.
This phenomenon may occur due to the behavior of the fsck command of some distributions (fsck accesses an unintended block device).

```
kernel: [I] <type: liscal><event: 144> NMPx I/O port has been closed,␣
↪mount(0), io(0).
kernel: [I] <type: liscal><event: 144> - This message can be recorded by fsck␣
↪command when NMPx becomes active.
kernel: [I] <type: liscal><event: 144> - This message can be recorded on␣
↪hotplug service starting when NMPx is not active.
kernel: [I] <type: liscal><event: 144> - Ignore this and following messages␣
↪'Buffer I/O error on device NMPx' on such environment.
kernel: Buffer I/O error on device /dev/NMPx, logical block xxxx

kernel: <liscal liscal_make_request> NMPx I/O port is close, mount(0), io(0).
kernel: Buffer I/O error on device /dev/NMPx , logical block xxxx
```

This is not a problem for EXPRESSCLUSTER. If this causes any problem such as heavy use of message files, change the following settings of mirror disk resources or hybrid disk resources.

- Select "Not Execute" on "fsck action before mount"
- Select "Execute" on "fsck Action When Mount Failed"

6.5.6 Messages displayed when loading a driver

When loading a mirror driver, messages like the following may be displayed at the console and/or syslog. However, this is not an error.

```
kernel: liscal: no version for "xxxxx" found: kernel tainted.
kernel: liscal: module license 'unspecified' taints kernel.
```

(Any character strings are set to *xxxxx* .)

And also, when loading the clpka or clpkhb driver, messages like the following may be displayed on the console and/or syslog. However, this is not an error.

```
kernel: clpkhb: no version for "xxxxx" found: kernel tainted.
kernel: clpkhb: module license 'unspecified' taints kernel.

kernel: clpka: no version for "xxxxx" found: kernel tainted.
kernel: clpka: module license 'unspecified' taints kernel.
```

(Any character strings are input into *xxxxx* .)

6.5.7 Messages displayed for the first I/O to mirror disk resources or hybrid disk resources

When reading/writing data from/to a mirror disk resource or hybrid disk resource for the first time after the resource was mounted, a message like the following may be displayed at the console and/or syslog. However, this is not an error.

```
kernel: JBD: barrier-based sync failed on NMPx - disabling barriers
```

(Any character strings are set to *x* .)

6.5.8 File operating utility on X-Window

Some of the file operating utilities (coping and moving files and directories via GUI) on X-Window perform the following:

- Checks if the block device is usable.
- Mounts the file system if there is any that can be mounted.

Make sure not to use file operating utility that perform above operations. They may cause problem to the operation of EXPRESSCLUSTER.

6.5.9 IPMI message

When you are using ipmi for user mode monitor resources, the following kernel module warning log is recorded many times in the syslog.

```
modprobe: modprobe: Can't locate module char-major-10-173
```

When you want to prevent this log from being recorded, rename /dev/ipmikcs.

6.5.10 Limitations during the recovery operation

Do not control the following commands, clusters and groups by the Cluster WebUI while recovery processing is changing (reactivation -> failover -> last operation), if a group resource is specified as a recovery target and when a monitor resource detects an error.

- Stop and suspend of a cluster
- Start, stop, moving of a group

If these operations are controlled at the transition to recovering due to an error detected by a monitor resource, the other group resources in the group may not be stopped.
Even if a monitor resource detects an error, it is possible to control the operations above after the last operation is performed.

6.5.11 Executable format file and script file not described in manuals

Executable format files and script files which are not described in "EXPRESSCLUSTER command reference" in the "Reference Guide" exist under the installation directory. Do not run these files on any system other than EXPRESSCLUSTER. The consequences of running these files will not be supported.

6.5.12 Executing fsck

- When fsck is specified to execute at activation of disk resources, mirror disk resources, or hybrid disk resources, fsck is executed when an ext2/ext3/ext4 file system is mounted. Executing fsck may take times depending on the size, usage or status of the file system, resulting that an fsck timeout occurs and mounting the file system fails.

  This is because fsck is executed in either of the following ways.

  (a) Only performing simplified journal check.
  Executing fsck does not take times.

  (b) Checking consistency of the entire file system.
  When the data saved by OS has not been checked for 180 days or more or the data will be checked after it is mounted around 30 times.
  In this case, executing fsck takes times depending the size or usage of the file system.
  Specify a time in safe for the fsck timeout of disk resources so that no timeout occurs.

- When fsck is specified not to execute at activation of disk resources, mirror disk resources, or hybrid disk resources, the warning described below may be displayed on the console and/or syslog when an ext2/ext3/ext4 file system is mounted more than the mount execution count set to OS that it is recommended to execute fsck.

  `EXT2-fs warning: xxxxx, running e2fsck is recommended.`

  Note: There are multiple patterns displayed in *xxxxx* .

  It is recommended to execute fsck when this waning is displayed.

  Follow the steps below to manually execute fsck.
  Be sure to execute the following steps on the server where the disk resource in question has been activated.

  (1) Deactivate a group to which the disk resource in question belongs by using a command such as clpgrp.

  (2) Confirm that no disks have been mounted by using a command such as mount and df.

  (3) Change the state of the disk from Read Only to Read Write by executing one of the following commands depending on the disk resource type.

  Example for disk resources: A device name is /dev/sbd5

  ```
  # clproset -w -d /dev/sbd5
  /dev/sbd5 : success
  ```

  Example for mirror disk resources: A resource name is md1.

  ```
  # clpmdctrl --active -nomount md1
  <md1@server1>: active successfully
  ```

  Example for hybrid disk resources: A resource name is hd1.

  ```
  # clphdctrl --active -nomount hd1
  <hd1@server1>: active successfully
  ```

(4) Execute fsck.
(If you specify the device name for fsck execution in the case of a mirror disk resource or hybrid disk resource, specify the mirror partition device name (/dev/NMP*x*) corresponding to the resource.)

(5) Change the state of the disk from Read Write to Read Only by executing one of the following commands depending on the disk resource type.

Example for disk resources: A device name is /dev/sbd5.

```
# clproset -o -d /dev/sdb5
/dev/sdb5 : success
```

Example for mirror disk resources: A resource name is md1.

```
# clpmdctrl --deactive md1
<md1@server1>: deactive successfully
```

Example for hybrid disk resources: A resource name is hd1.

```
# clphdctrl --deactive hd1
<hd1@server1>: deactive successfully
```

(6) Activate a group to which the disk resource in question belongs by using a command such as clpgrp.

If you need to specify that the warning message is not output without executing fsck, for ext2/ext3/ext4, change the maximum mount count by using tune2fs. Be sure to execute this command on the server where the disk resource in question has been activated.

(1) Execute one of the following commands..

Example for disk resources: A device name is /dev/sbd5.

```
# tune2fs -c -1 /dev/sdb5
tune2fs 1.42.9 (28-Dec-2013)
Setting maximal mount count to -1
```

Example for mirror disk resources: A mirror partition device name is /dev/NMP1.

```
# tune2fs -c -1 /dev/NMP1
tune2fs 1.42.9 (28-Dec-2013)
Setting maximal mount count to -1
```

Example for hybrid disk resources: A mirror partition device name is /dev/NMP1.

```
# tune2fs -c -1 /dev/NMP1
tune2fs 1.42.9 (28-Dec-2013)
Setting maximal mount count to -1
```

(2) Confirm that the maximum mount count has been changed.

Example: A device name is /dev/sbd5.

```
# tune2fs -l /dev/sdb5
tune2fs 1.42.9 (28-Dec-2013)
Filesystem volume name: <none>
    :
Maximum mount count: -1
    :
```

6.5.13 Executing xfs_repair

When an xfs-based disk resource/mirror disk resource/hybrid disk resource is activated, the console may display a warning message of xfs. In this case, executing xfs_repair is recommended to restore the file system.

To run xfs_repiar, follow these steps:

1. Make sure that the resource is not activated. If the resource is activated, deactivate it with Cluster WebUI.
2. Make the device writable.

   Example of a disk resource whose device name is /dev/sdb1:

   ```
   # clproset -w -d /dev/sdb1
   /dev/sdb1 : success
   ```

   Example of a mirror disk resource whose name is md1:

   ```
   # clpmdctrl --active -nomount md1
   <md1@server1>: active successfully
   ```

   Example of a hybrid disk resource whose name is hd1:

   ```
   # clphdctrl --active -nomount hd1
   <hd1@server1>: active successfully
   ```

3. Mount the device.

   Example of a disk resource whose device name is /dev/sdb1:

   ```
   # mount /dev/sdb1 /mnt
   ```

   Example of a mirror/hybrid disk resource whose mirror partition device name is /dev/NMP1:

   ```
   # mount /dev/NMP1 /mnt
   ```

4. Unmount the device.

   ```
   # umount /mnt
   ```

   Note: The xfs_repair utility cannot restore a file system including a dirty log. Such a file system need be mounted and then unmounted to clear the log.

5. Execute xfs_repair.

   Example of a disk resource whose device name is /dev/sdb1:

   ```
   # xfs_repair /dev/sdb1
   ```

   Example of a mirror/hybrid disk resource whose mirror partition device name is /dev/NMP1:

   ```
   # xfs_repair /dev/NMP1
   ```

6. Write-protect the device.

   Example of a disk resource whose device name is /dev/sdb1:

   ```
   # clproset -o -d /dev/sdb1
   /dev/sdb1 : success
   ```

   Example of a mirror disk resource whose name is md1:

   ```
   # clpmdctrl --deactive md1
   <md1@server1>: deactive successfully
   ```

   Example of a hybrid disk resource whose name is hd1:

```
# clphdctrl --deactive hd1
<hd1@server1>: deactive successfully
```

Now you have finished restoring the xfs file system.

6.5.14 Messages when collecting logs

When collecting logs, the message described below is displayed at the console, but this is not an error. Logs are collected successfully.

```
hd#: bad special flag: 0x03
ip_tables: (C) 2000-2002 Netfilter core team
```

("*hd#*" is replaced with the device name of IDE.)

```
kernel: Warning: /proc/ide/hd?/settings interface is obsolete, and will be␣
↪removed soon!
```

6.5.15 Failover and activation during mirror recovery

- When mirror recovery is in progress for a mirror disk resource or hybrid disk resource, a mirror disk resource or hybrid disk resource placed in the deactivated state cannot be activated.

  During mirror recovery, a failover group including the disk resource in question cannot be moved.

  If a failover occurs during mirror recovery, the copy destination server does not have the latest status, so a failover to the copy destination server or copy destination server group will fail.

  Even if an attempt to fail over a hybrid disk resource to a server in the same server group is made by actions for when a monitor resource detects an error, it will fail, too, since the current server is not changed.

  Note that, depending on the timing, when mirror recovery is completed during a failover, move, or activation, the operation may be successful.

- At the first mirror startup after configuration information registration and also at the first mirror startup after a mirror disk is replaced after a failure, the initial mirror configuration is performed.

  In the initial mirror configuration, disk copying (full mirror recovery) is performed from the active server to the mirror disk on the standby server immediately after mirror activation.

  Until this initial mirror configuration (full mirror recovery) is completed and the mirror enters the normal synchronization state, do not perform either failover to the standby server or group movement to the standby server.

  If a failover or group movement is performed during this disk copying, the standby server may be activated while the mirror disk of the standby server is still incomplete, causing the data that has not yet been copied to the standby server to be lost and thus causing mismatches to occur in the file system.

6.5.16 Cluster shutdown and reboot (mirror disk resource and hybrid disk resource)

When using a mirror disk resource or a hybrid disk resource, do not execute cluster shutdown or cluster shutdown reboot from the clpstdn command or the Cluster WebUI while a group is being activated.

A group cannot be deactivated while a group is being activated. Therefore, OS may be shut down in the state that mirror disk resource or hybrid disk resources is not deactivated successfully and a mirror break may occur.

6.5.17 Shutdown and reboot of individual server (mirror disk resource and hybrid disk resource)

When using a mirror disk and a hybrid disk resource, do not shut down the server or run the shutdown reboot command from the clpdown command or the Cluster WebUI while activating the group.
A group cannot be deactivated while a group is being activated. Therefore, OS may be shut down and a mirror break may occur in the state that mirror disk resources and hybrid disk resources are not deactivated successfully.

6.5.18 Scripts for starting/stopping EXPRESSCLUSTER services

For an init.d environment, an error occurs in the service startup and stop scripts in the following cases. For a systemd environment, an error does not occur.

- Before start operating EXPRESSCLUSTER
  When a server start up, the error occurs in the following starting scripts. There is no problem for the error because cluster configuration data has not uploaded.
  - clusterpro_md
- At following case, the script to terminate EXPRESSCLUSTER services may be executed in the wrong order.
  The OS is shut down after EXPRESSCLUSTER services are disabled.
  EXPRESSCLUSTER services may be terminated in the wrong order at OS shutdown if all of EXPRESSCLUSTER services are disabled. This problem is caused by failure in termination process for the service has been already disabled.
  As long as the system shutdown is executed by Cluster WebUI or clpstdn command, there is no problem even if the services is terminated in the wrong order. But, any other problem may not be happened by wrong order termination.

6.5.19 Service startup time

EXPRESSCLUSTER services might take a while to start up, depending on the wait processing at startup.

- clusterpro_evt
  Servers other than the master server wait up to two minutes for configuration data to be downloaded from the master server. Downloading usually finishes within several seconds if the master server is already operating. The master server does not have this wait process.
- clusterpro_nm
  There is no wait process. This process usually finishes within several seconds.
- clusterpro_trn
  There is no wait process. This process usually finishes within several seconds.
- clusterpro_ib
  There is no wait process. This process usually finishes within several seconds.
- clusterpro_api
  There is no wait process. This process usually finishes within several seconds.
- clusterpro_md
  This service starts up only when the mirror or hybrid disk resources exist. The system waits up to one minute for the mirror agent to normally start up. This process usually finishes within several seconds.
- clusterpro

Although there is no wait process, EXPRESSCLUSTER might take several tens of seconds to start up. This process usually finishes within several seconds.

- clusterpro_webmgr
  There is no wait process. This process usually finishes within several seconds.
- clusterpro_alertsync
  There is no wait process. This process usually finishes within several seconds.

In addition, the system waits for cluster activation synchronization after the EXPRESSCLUSTER daemon is started. By default, this wait time is five minutes.
For details, see "The system maintenance information" in the "Maintenance Guide".

6.5.20 Checking the service status in a systemd environment

For a systemd environment, the service status displayed by using the systemctl command may differ from the actual cluster status.
Use the clpstat command or Cluster WebUI to check the cluster status.

6.5.21 Scripts in EXEC resources

EXEC resource scripts of group resources stored in the following location.

*/opt/nec/clusterpro/*scripts/*group-name/resource-name/*

The following cases, old EXEC resource scripts are not deleted automatically.

- When the EXEC resource is deleted or renamed
- When a group that belongs to the EXEC resource is deleted or renamed

Old EXEC resource scripts can be deleted when unnecessary.

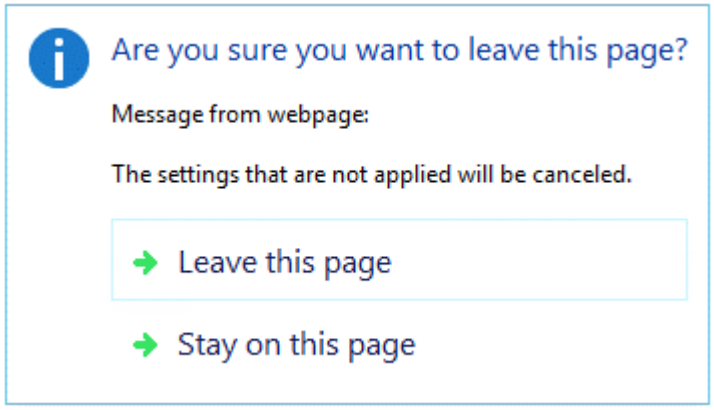
6.5.22 Monitor resources that monitoring timing is "Active"

When monitor resources that monitoring timing is "Active" have suspended and resumed, the following restriction apply:

- In case stopping target resource after suspending monitor resource, monitor resource becomes suspended. As a result, monitoring restart cannot be executed.
- In case stopping or starting target resource after suspending monitor resource, monitoring by monitor resource starts when target resource starts.

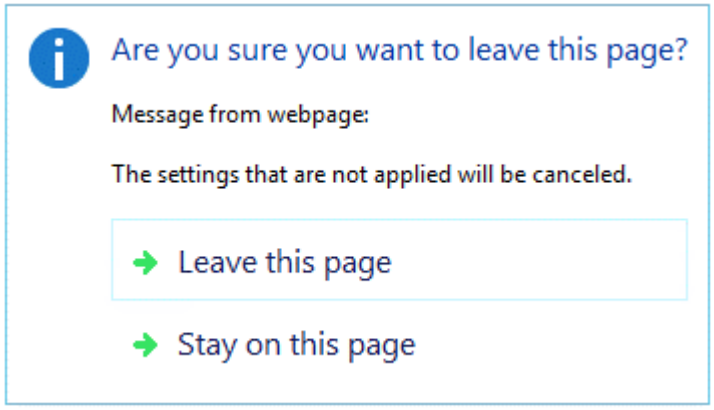
6.5.23 Notes on the Cluster WebUI

- If the Cluster WebUI is operated in the state that it cannot communicate with the connection destination, it may take a while until the control returns.
- When going through the proxy server, make the settings for the proxy server be able to relay the port number of the Cluster WebUI.
- When going through the reverse proxy server, the Cluster WebUI will not operate properly.
- When updating EXPRESSCLUSTER, close all running browsers. Clear the browser cache and restart the browser.
- Cluster configuration data created using a later version of this product cannot be used with this product.
- When closing the Web browser, the dialog box to confirm to save may be displayed.



When you continue to edit, click the **Stay on this page**button.

- Reloading the Web browser (by selecting **Refresh button** from the menu or tool bar), the dialog box to confirm to save may be displayed.



When you continue to edit, click the **Stay on this page**button.

- For notes and restrictions of Cluster WebUI other than the above, see the online manual.

6.5.24 Changing the partition size of mirror disks and hybrid disk resources

When changing the size of mirror partitions after the operation is started, see "Changing offset or size of a partition on mirror disk resource" in "The system maintenance information" in the "Maintenance Guide".

6.5.25 Changing kernel dump settings

- If you are changing the kdump settings and "applying" them through "kernel dump configuration" (system-config-kdump) while the cluster is running on Red Hat Enterprise Linux 6 or the like, you may see the following error message output.

  In this case, stop the cluster once (stop the mirror agent as well as the cluster when using a mirror disk resource or hybrid disk resource), and then retry the kernel dump configuration.

  * The following *{driver_name}* indicates clpka, clpkhb, or liscal.

  ```
  No module {driver_name} found for kernel {kernel_version}, aborting
  ```

6.5.26 Notes on floating IP and virtual IP resources

- Do not execute a network restart on a server on which floating IP resources or virtual IP resources are active. If the network is restarted, any IP addresses that have been added as floating IP resources or virtual IP resources are deleted.

6.5.27 System monitor resources,Process resource monitor resource

- To change a setting, the cluster must be suspended.
- System monitor resources do not support a delay warning for monitor resources.
- If the date and time of the OS is changed during operation, the timing of analysis processing being performed at 10-minute intervals will change only once immediately after the date and time is changed. This will cause the following to occur; suspend and resume the cluster as necessary.
  - An error is not detected even when the time to be detected as abnormal elapses.
  - An error is detected before the time to be detected as abnormal elapses.
  - Up to 64 disks can be monitored at the same time by the disk resource monitor function of system monitor resource.

6.5.28 JVM monitor resources

- When restarting the monitoring-target Java VM, suspend or shut down the cluster before restarting the Java VM.
- To change a setting, the cluster must be suspended.
- JVM monitor resources do not support a delay warning for monitor resources.

6.5.29 HTTP monitor resource

- The HTTP monitor resource uses any of the following OpenSSL shared library symbolic links:
  - libssl.so
  - libssl.so.1.1 (OpenSSL 1.1.1 shared libraly)
  - libssl.so.10 (OpenSSL 1.0 shared libraly)
  - libssl.so.6 (OpenSSL 0.9 shared libraly)

The above symbolic links may not exist depending on the OS distribution or version, or the package installation status.
If the above symbolic links cannot be found, the following error occurs in the HTTP monitor resource.

```
Detected an error in monitoring<Module Resource Name>. (1 :Can not found␣
↪library. (libpath=libssl.so, errno=2))
```

For this reason, if the above error occurred, be sure to check whether the above symbolic links exit in /usr/lib or /usr/lib64.
If the above symbolic links do not exit, create the symbolic link libssl.so, as in the command example below.

Command example:

```
cd /usr/lib64                          # Move to /usr/lib64.
ln -s libssl.so.1.0.1e libssl.so       # Create a symbolic link.
```

6.5.30 Restoration from an AMI in an AWS environment

- If the ENI ID of a primary network interface is set to the **ENI ID** of the AWS virtual ip resource and AWS elastic ip resource and AWS secondary ip resource, the AWS virtual ip resource and AWS elastic ip resource and AWS secondary ip resource setting is required to change when restoring data from an AMI.

If the ENI ID of a secondary network interface is set to the **ENI ID** of the AWS virtual ip resource and AWS elastic ip resource and AWS secondary ip resource, it is unnecessary to set the AWS virtual ip resource and AWS elastic ip resource and AWS secondary ip resource again because the same ENI ID is inherited by a detach/attach processing when restoring data from an AMI.

6.6 Notes when changing the EXPRESSCLUSTER configuration

The section describes what happens when the configuration is changed after starting to use EXPRESSCLUSTER in the cluster configuration.

6.6.1 Exclusive rule of group properties

When the exclusive attribute of the exclusive rule is changed, the change is applied by suspending and resuming the cluster.
When a group is added to the exclusion rule whose exclusive attribute is set to Absolute, multiple groups of **Absolute** exclusive start on the same server depending on the group startup status before suspending the cluster.
Exclusive control will be performed at the next group startup.

6.6.2 Dependency between resource properties

When the dependency between resources has been changed, the change is applied by suspending and resuming the cluster.
If a change in the dependency between resources that requires the resources to be stopped during application is made, the startup status of the resources after the resume may not reflect the changed dependency.
Dependency control will be performed at the next group startup.

6.6.3 Deleting disk resources

When a disk resource is deleted, the corresponding device is sometimes set to **Read Only**.

Change the status of the device to **Read Write** by using the clproset command.

6.6.4 Setting cluster statistics information of message receive monitor resources

Once the settings of cluster statistics information of monitor resource has been changed, the settings of cluster statistics information are not applied to message receive monitor resources even if you execute the suspend and resume. Reboot the OS to apply the settings to the message receive monitor resources.

6.6.5 Changing a port number

If you have changed a port number with the server firewall enabled, the firewall configuration needs to be changed as well by using the clpfwctrl command. For more information, see "Reference Guide" -> "EXPRESSCLUSTER command reference" -> "Adding a firewall rule (clpfwctrl command)".

6.7 Notes on upgrading EXPRESSCLUSTER

This section describes the notes on upgrading EXPRESSCLUSTER after starting a cluster operation.

6.7.1 Changed Functions

The following describes the functions changed for each of the versions.

Internal Version 4.0.0-1

- Management tool
  The default management tool has been changed to Cluster WebUI. If you want to use the conventional WebManager as the management tool, specify "http://management IP address of management group or actual IP address:port number of the server in which EXPRESSCLUSTER Server is installed/main.htm" in the address bar of a web browser.
- Mirror/hybrid disk resource
  Considering that the minimum size of a cluster partition has been increased to 1 GiB, prepare a sufficient size of it for upgrading EXPRESSCLUSTER.

Internal Version 4.1.0-1

- Configuration tool
  The default configuration tool has been changed to Cluster WebUI, which allows you to manage and configure clusters with Cluster WebUI.
- Cluster statistical information collection function
  By default, the cluster statistical information collection function saves statistics information files under the installation path. To avoid saving the files for such reasons as insufficient disk capacity, disable the cluster statistical information collection function. For more information on settings for this function, see "Parameter details" in the "Reference Guide".
- Mirror/hybrid disk resource in the asynchronous mode
  In the asynchronous mode, the mirror break status is not set even if the queue for the data to be sent has become full. The overflown ones are temporarily written as a history file. Due to this functional enhancement, it is necessary to enter the setting values below:
  - History file storage directory
  - History file size limit

  * Shortly after the update, these setting values are in blank. In this case, the history file storage directory is treated as a directory which has installed EXPRESSCLUSTER, and no limit is imposed on the history file size.

  For more information on the setting values, see "Understanding mirror disk connect monitor resources" of "Group resource details" in the "Reference Guide".

- System monitor resource

  The **System Resource Agent process settings** part of the system monitor resource has been separated to become a new monitor resource. Therefore, the conventional monitor settings of the **System Resource Agent process settings** are no longer valid. To continue the conventional monitoring, configure it by registering a new

process resource monitor resource after upgrading EXPRESSCLUSTER. For more information on monitor settings for process resource monitor resources, see "Understanding process resource monitor resources" in "Monitor resource details" in the "Reference Guide".

Internal Version 4.2.0-1

- AWS AZ monitor resource
  The way of evaluating the AZ status grasped through the AWS CLI has been changed: available as normal, information or impaired as warning, and unavailable as warning. (Previously, any AZ status other than **available** was evaluated as abnormal.)

Internal Version 4.3.0-1

- Weblogic monitor resource
  REST API has been added as a new monitoring method. From this version, REST API is the default value for the monitoring method. At the version upgrade, reconfigure the monitoring method.
  The default value of the password has been changed. If you use weblogic that is the previous default value, reset the password default value.

Internal Version 5.0.0-1

- Forced stop function and scripts
  These have been redesigned as individual forced stop resources adapted to environment types.
  Since the forced stop function and scripts configured before the upgrade are no longer effective, set them up again as forced stop resources.

6.7.2 Removed Functions

The following describes the functions removed for each of the versions.

Internal Version 4.0.0-1

- WebManager Mobile
- OracleAS monitor resource

Important:

Upgrading EXPRESSCLUSTER from its old version requires manually updating the cluster configuration data for functions with corresponding actions described in the table below.

For information on how to upgrade EXPRESSCLUSTER, see "7.1.1. *How to upgrade from X3.3 or X4.x to X 5.0*". Then, at the timing described in the guide, follow each of the procedures described in the Action column.

Internal Version 5.0.0-1

Function	Action
WebManager/Builder	
COM heartbeat resource	1. Open **Cluster Properties** -> **Interconnect tab**, then remove each heartbeat interface whose type is **unknown**.

Continued on next page

Table 6.17 – continued from previous page

Function	Action
Virtual machine groups Virtual machine resources Virtual machine monitor resources	You cannot move configuration data (for a host cluster) which involves virtual machine groups.
BMC linkage	1. Delete relevant message reception monitor resources.
Controlling CPU frequency command (clpcpufreq command)	-
Estimating the amount of resource usage command (clpprer command)	-
Controlling chassis identify lamp command (clpledctrl command)	-
Processing inter-cluster linkage command (clptrnreq command)	-
Changing BMC information command (clpbmccnf command)	-
Disk I/O Lockout	-
Disk Heart Beat RawDevice	-
IBM POWER does not support the following functions: kernel mode LAN heartbeat resources User mode monitor resources - Monitor by: keepalive Keepalive Reset Keepalive Panic	It does not apply to IBM POWER and IBM POWER LE

Continued on next page

Table 6.17 – continued from previous page

Function	Action
NAS resources	1. If NAS resources are individually set in group resources' dependency, remove the dependency settings first. For the group resources, open **Resource Properties** -> the **Dependency** tab, select the NAS resources, and then click the **deleted** button to exclude them from the dependency. 2. Delete NAS resources.
Load Balancer Linkage Settings (JVM monitor resources)	-
Sybase monitor resources	-
Controlling the rest point of Sybase command (clpsybasestill command)	-
VxVM linkage Disk resource - Disk Type(VXVM) Volume manager resource - Volume manager(VXVM) Disk monitor resource - Method (READ (VXVM)) Volume manager monitor resource - Volume manager(VXVM)	Configuration data based on the VxVM linkage function cannot be migrated.

6.7.3 Removed Parameters

The following tables show the parameters configurable with Cluster WebUI but removed for each of the versions.

Internal Version 4.0.0-1

Cluster

Parameters	default values
Cluster Properties	
Alert Service Tab	
• Use Alert Extension	Off
WebManager Tab	
• Enable WebManager Mobile Connection	Off

Continued on next page

Table 6.18 – continued from previous page

Parameters	default values
Web Manager Mobile Password	
• Password for Operation	-
• Password for Reference	-

JVM monitor resource

Parameters	default values
JVM Monitor Resource Properties	
Monitor(special) Tab	
Memory Tab (when Oracle Java is selected for JVM Type)	
• Monitor Virtual Memory Usage	2048 MB
Memory Tab (when Oracle JRockit is selected for JVM Type)	
• Monitor Virtual Memory Usage	2048 MB
Memory Tab(when Oracle Java(usage monitoring) is selected for JVM Type)	
• Monitor Virtual Memory Usage	2048 MB

Internal Version 4.1.0-1

Cluster

Parameters	default values
Cluster Properties	
WebManager Tab	
WebManager Tuning Properties	
Behavior Tab	
• Max. Number of Alert Records on the Viewer	300
• Client Data Update Method	Real Time

Internal Version 5.0.0-1

Cluster

Parameters	default values
Cluster Properties	
Interconnect Tab	
COM Device	
Disk Heart Beat Properties	
RawDevice	
Extension Tab	
• Virtual Machine Forced Stop Setting - Virtual Machine Management Tool	vCenter
• Virtual Machine Forced Stop Setting - Command	/usr/lib/vmware-vcli/apps/vm/vmcontrol.pl
• Execute Script for Forced Stop	Off
Server Properties	
Info Tab	
• Virtual Machine	Off
• Type	vSphere
BMC Tab	
• Forced Stop Command Line	-
• Chassis Identify - Flash / Turn off	-
Disk I/O Lockout Tab	
• I/F No. (Add, Remove)	The order you added I/Fs
• Device (Edit)	-

6.7.4 Changed Default Values

The following tables show the parameters which are configurable with Cluster WebUI but whose defaults have been changed for each of the versions.

- To continue using a "Default value before update" after the upgrade, change the corresponding "Default value after update" to the desired one.
- Any setting other than a "Default value before update" is inherited to the upgraded version and therefore does not need to be restored.

Internal Version 4.0.0-1

Cluster

Parameters	Default value before update	Default value after update
Cluster Properties		
Monitor Tab		
• Method	softdog	keepalive
JVM monitor Tab		
• Maximum Java Heap Size	7 MB	16 MB

Exec resource

Parameters	Default value before update	Default value after update
Exec Resource Properties		
Dependence Tab		
• Follow the default dependence	On - floating IP resources - virtual IP resources - disk resources - mirror disk resources - hybrid disk resources - NAS resources - Dynamic DNS resource - Volume manager resource - AWS elastic ip resource - AWS virtual ip resource - Azure probe port resource	On - floating IP resources - virtual IP resources - disk resources - mirror disk resources - hybrid disk resources - NAS resources - Dynamic DNS resource - Volume manager resource - AWS elastic ip resource - AWS virtual ip resource - AWS DNS resource - Azure probe port resource - Azure DNS resource

Disk resource

Parameters	Default value before update	Default value after update
Disk Resource Properties		
Dependence Tab		

Continued on next page

Table 6.24 – continued from previous page

Parameters	Default value before update	Default value after update
• Follow the default dependence	On - floating IP resources - virtual IP resources - Dynamic DNS resource - Volume manager resource - AWS elastic ip resource - AWS virtual ip resource - Azure probe port resource	On - floating IP resources - virtual IP resources - Dynamic DNS resource - Volume manager resource - AWS elastic ip resource - AWS virtual ip resource - AWS DNS resource - Azure probe port resource - Azure DNS resource
Details Tab		
Disk Resource Tuning Properties		
Mount Tab		
• Timeout	60 sec	180 sec
xfs_repair Tab (when xfs is selected for File System)		
• xfs_repair Action When Mount Failed Execute	On	Off

NAS resource

Parameters	Default value before update	Default value after update
NAS Resource Properties		
Dependence Tab		
• Follow the default dependence	On - floating IP resources - virtual IP resources - Dynamic DNS resources - AWS elastic ip resource - AWS virtual ip resource - Azure probe port resource	On - floating IP resources - virtual IP resources - Dynamic DNS resources - AWS elastic ip resource - AWS virtual ip resource - AWS DNS resource - Azure probe port resource - Azure DNS resource

Mirror disk resource

Parameters	Default value before update	Default value after update
Mirror Disk Resource Properties		
Dependency Tab		
• Follow the default dependence	On - floating IP resources - virtual IP resources - AWS elastic ip resource - AWS virtual ip resource - Azure probe port resource	On - floating IP resources - virtual IP resources - AWS elastic ip resource - AWS virtual ip resource - AWS DNS resource - Azure probe port resource - Azure DNS resource
Details Tab		
Mirror Disk Resource Tuning Properties		
xfs_repair Tab (when xfs is selected for File System)		
• xfs_repair Action When Mount Failed Execute	On	Off

Hybrid disk resource

Parameters	Default value before	Default value after update
Hybrid Disk Resource Properties		
Dependency Tab		
• Follow the default dependence	On - floating IP resources - virtual IP resources - AWS elastic ip resource - AWS virtual ip resource - Azure probe port resource	On - floating IP resources - virtual IP resources - AWS elastic ip resource - AWS virtual ip resource - AWS DNS resource - Azure probe port resource - Azure DNS resource
Details Tab		
Hybrid Disk Resource Tuning Properties		
xfs_repair Tab (when xfs is selected for File System)		

Continued on next page

Table 6.27 – continued from previous page

Parameters	Default value before	Default value after update
• xfs_repair Action When Mount Failed Execute	On	Off

Volume manager resource

Parameters	Default value before update	Default value after update
Volume Manager Resource Properties		
Dependency Tab		
• Follow the default dependence	On - Floating IP resources - Virtual IP resources - Dynamic DNS resources - AWS elastic ip resource - AWS virtual ip resource - Azure probe port resource	On - Floating IP resources - Virtual IP resources - Dynamic DNS resources - AWS elastic ip resource - AWS virtual ip resource - AWS DNS resource - Azure probe port resource - Azure DNS resource

Virtual IP monitor resource

Parameters	Default value before update	Default value after update
Virtual IP Monitor Resource Properties		
Monitor(common)		
• Timeout	30 sec	180 sec

PID monitor resource

Parameters	Default value before update	Default value after update
PID Monitor Resource Properties		
Monitor(common)Tab		
• Wait Time to Start Monitoring	0 sec	3 sec

Continued on next page

Table 6.30 – continued from previous page

Parameters	Default value before update	Default value after update
• Do Not Retry at Timeout Occurrence	Off	On
• Do not Execute Recovery Action at Timeout Occurrence	Off	On

User mode monitor resource

Parameters	Default valu before update	Default value after update
User mode Monitor Resource Properties		
Monitor(special) Tab		
• Method	softdog	keepalive

NIC Link Up/Down monitor resource

Parameters	Default value before update	Default value after update
NIC Link Up/Down Monitor Resource Properties		
Monitor(common) Tab		
• Timeout	60 sec	180 sec
• Do Not Retry at Timeout Occurrence	Off	On
• Do not Execute Recovery Action at Timeout Occurrence	Off	On

ARP monitor resource

Parameters	Default value before update	Default value after update
ARP Monitor Resource Properties		
Monitor(common) Tab		

Continued on next page

Table 6.33 – continued from previous page

Parameters	Default value before update	Default value after update
• Do Not Retry at Timeout Occurrence	Off	On
• Do not Execute Recovery Action at Timeout Occurrence	Off	On

Dynamic DNS monitor resource

Parameters	Default value before update	Default value after update
Dynamic DNS Monitor Resource Properties		
Monitor(common) Tab		
• Timeout	100 sec	180 sec

Process name monitor resource

Parameters	Default value before update	Default value after update
Process Monitor Resource Properties		
Monitor(common) tab		
• Wait Time to Start Monitoring	0 sec	3 sec
• Do Not Retry at Timeout Occurrence	Off	On
• Do not Execute Recovery Action at Timeout Occurrence	Off	On

DB2 monitor resource

Parameters	Default value before update	Default value after update
DB2 Monitor Resource Properties		

Continued on next page

Table 6.36 – continued from previous page

Parameters	Default value before update	Default value after update
Monitor(special) Tab		
• Password	ibmdb2	-
• Library Path	`/opt/IBM/db2/V8.2/lib/libdb2.so`	`/opt/ibm/db2/V11.1/lib64/libdb2.so`

MySQL monitor resource

Parameters	Default value before update	Default value after update
MySQL Monitor Resource Properties		
Monitor(special) Tab		
• Storage Engine	MyISAM	InnoDB
• Library Path	`/usr/lib/mysql/libmysqlclient.so.15`	`/usr/lib64/mysql/libmysqlclient.so.20`

Oracle monitor resource

Parameters	Default value before update	Default value after update
Oracle Monitor Resource Properties		
Monitor(special) Tab		
• Password	change_on_install	-
• Library Path	`/opt/app/oracle/product/10.2.0/db_1/lib/libclntsh.so.10.1`	`/u01/app/oracle/product/12.2.0/dbhome_1/lib/libclntsh.so.12.1`

PostgreSQL monitor resource

Parameters	Default value before update	Default value after update
PostgreSQL Monitor Resource Properties		
Monitor(special) Tab		
• Library Path	`/usr/lib/libpq.so.3.0`	`/opt/PostgreSQL/10/lib/libpq.so.5.10`

Tuxedo monitor resource

Parameters	Default value before update	Default value after update
Tuxedo Monitor Resource Properties		
Monitor(special) Tab		
• Library Path	`/opt/bea/tuxedo8.1/lib/libtux.so`	`/home/Oracle/tuxedo/tuxedo12.1.3.0.0/lib/libtux.so`

Weblogic monitor resource

Parameters	Default value before update	Default value after update
Weblogic Monitor Resource Properties		
Monitor(special) Tab		
• Domain Environment File	`/opt/bea/weblogic81/samples/domains/examples/setExamplesEnv.sh`	`/home/Oracle/product/Oracle_Home/user_projects/domains/base_domain/bin/setDomainEnv.sh`

JVM monitor resource

Parameters	Default value before update	Default value after update
JVM Monitor Resource Properties		
Monitor(common) Tab		
• Timeout	120 sec	180 sec

Floating IP monitor resources

Parameters	Default value before update	Default value after update
Floating IP Monitor Resource Properties		
Monitor(common) Tab		
• Timeout	60 sec	180 sec
• Do Not Retry at Timeout Occurrence	Off	On

Continued on next page

Table 6.43 – continued from previous page

Parameters	Default value before update	Default value after update
• Do not Execute Recovery Action at Timeout Occurrence	Off	On

AWS Elastic IP monitor resource

Parameters	Default value before update	Default value after update
AWS elastic ip Monitor Resource Properties		
Monitor(common) Tab		
• Timeout	100 sec	180 sec
• Do Not Retry at Timeout Occurrence	Off	On
• Do not Execute Recovery Action at Timeout Occurrence	Off	On

AWS Virtual IP monitor resource

Parameters	Default value before update	Default value after update
AWS virtual ip Monitor Resource Properties		
Monitor(common) Tab		
• Timeout	100 sec	180 sec
• Do Not Retry at Timeout Occurrence	Off	On
• Do not Execute Recovery Action at Timeout Occurrence	Off	On

AWS AZ monitor resource

Parameters	Default value before update	Default value after update
AWS AZ Monitor Resource Properties		
Monitor(common) Tab		
• Timeout	100 sec	180 sec
• Do Not Retry at Timeout Occurrence	Off	On
• Do not Execute Recovery Action at Timeout Occurrence	Off	On

Azure probe port monitor resource

Parameters	Default value before update	Default value after update
Azure probe port Monitor Resource Properties		
Monitor(common) Tab		
• Timeout	100 sec	180 sec
• Do Not Retry at Timeout Occurrence	Off	On
• Do not Execute Recovery Action at Timeout Occurrence	Off	On

Azure load balance monitor resource

Parameters	Default value before update	Default value after update
Azure load balance monitor resource Properties		
Monitor(common) Tab		
• Timeout	100 sec	180 sec

Continued on next page

Table 6.48 – continued from previous page

Parameters	Default value before update	Default value after update
• Do Not Retry at Timeout Occurrence	Off	On
• Do not Execute Recovery Action at Timeout Occurrence	Off	On

Internal Version 4.1.0-1

Cluster

Parameters	Default value before update	Default value after update
Cluster Properties		
Monitor Tab		
• Shutdown monitor	Always execute	Execute when the group deactivation has been failed

Internal Version 4.2.0-1

AWS Elastic IP monitor resource

Parameters	Default value before update	Default value after update
AWS elastic ip Monitor Resource Properties		
Monitor(special) Tab		
• Action when AWS CLI command failed to receive response	Disable recovery action(Display warning)	Disable recovery action(Do nothing)

AWS Virtual IP monitor resource

Parameters	Default value before update	Default value after update
AWS virtual ip Monitor Resource Properties		
Monitor(special) Tab		
• Action when AWS CLI command failed to receive response	Disable recovery action(Display warning)	Disable recovery action(Do nothing)

AWS AZ monitor resource

Parameters	Default value before update	Default value after update
AWS AZ Monitor Resource Properties		
Monitor(special) Tab		
• Action when AWS CLI command failed to receive response	Disable recovery action(Display warning)	Disable recovery action(Do nothing)

AWS DNS monitor resource

Parameters	Default value before update	Default value after update
AWS DNS Monitor Resource Properties		
Monitor(special) Tab		
• Action when AWS CLI command failed to receive response	Disable recovery action(Display warning)	Disable recovery action(Do nothing)

Internal Version 4.3.0-1

Cluster

Parameters	Default value before update	Default value after update
Cluster Properties		
Extension Tab		
• Max Reboot Count	0 times	3 times
• Max Reboot Count Reset Time	0 min	60 min

NFS monitor resource

Parameters	Default value before update	Default value after update
NFS Monitor Resource Properties		
Monitor(special) Tab		

Continued on next page

Table 6.55 – continued from previous page

Parameters	Default value before update	Default value after update
• NFS Version	v2	v4

Weblogic monitor resource

Parameters	Default value before update	Default value after update
Weblogic Monitor Resource Properties		
Monitor(special) Tab		
• Password	weblogic	None

Internal Version 5.0.0-1

Cluster

Parameters	Default value before update	Default value after update
Cluster Properties		
Monitor Tab		
• Enable SIGTERM handler	Off	On

Exec resource

Parameters	Default value before update	Default value after update
Exec Resource Properties		
Dependence Tab		

Continued on next page

Table 6.58 – continued from previous page

Parameters	Default value before update	Default value after update
• Follow the default dependence	On - floating IP resources - virtual IP resources - disk resources - mirror disk resources - hybrid disk resources - Dynamic DNS resource - Volume manager resource - AWS elastic ip resource - AWS virtual ip resource - AWS DNS resource - Azure probe port resource - Azure DNS resource	On - floating IP resources - virtual IP resources - disk resources - mirror disk resources - hybrid disk resources - Dynamic DNS resource - Volume manager resource - AWS elastic ip resource - AWS virtual ip resource - AWS secondary ip resource - AWS DNS resource - Azure probe port resource - Azure DNS resource

Disk resource

Parameters	Default value before update	Default value after update
Disk Resource Properties		
Dependence Tab		
• Follow the default dependence	On - floating IP resources - virtual IP resources - Dynamic DNS resource - Volume manager resource - AWS elastic ip resource - AWS virtual ip resource - AWS DNS resource - Azure probe port resource - Azure DNS resource	On - floating IP resources - virtual IP resources - Dynamic DNS resource - Volume manager resource - AWS elastic ip resource - AWS virtual ip resource - AWS secondary ip resource - AWS DNS resource - Azure probe port resource - Azure DNS resource

Mirror disk resource

Parameters	Default value before update	Default value after update
Mirror Disk Resource Properties		

Continued on next page

Table 6.60 – continued from previous page

Parameters	Default value before update	Default value after update
Dependency Tab		
• Follow the default dependence	On - floating IP resources - virtual IP resources - AWS elastic ip resource - AWS virtual ip resource - Azure probe port resource	On - floating IP resources - virtual IP resources - AWS elastic ip resource - AWS virtual ip resource - AWS secondary ip resource - Azure probe port resource

Hybrid disk resource

Parameters	Default value before	Default value after update
Hybrid Disk Resource Properties		
Dependency Tab		
• Follow the default dependence	On - floating IP resources - virtual IP resources - AWS elastic ip resource - AWS virtual ip resource - Azure probe port resource	On - floating IP resources - virtual IP resources - AWS elastic ip resource - AWS virtual ip resource - AWS secondary ip resource - Azure probe port resource

Volume manager resource

Parameters	Default value before update	Default value after update
Volume Manager Resource Properties		
Dependency Tab		

Continued on next page

Table 6.62 – continued from previous page

Parameters	Default value before update	Default value after update
• Follow the default dependence	On - Floating IP resources - Virtual IP resources - Dynamic DNS resources - AWS elastic ip resource - AWS virtual ip resource - AWS DNS resource - Azure probe port resource - Azure DNS resource	On - Floating IP resources - Virtual IP resources - Dynamic DNS resources - AWS elastic ip resource - AWS virtual ip resource - AWS secondary ip resource - AWS DNS resource - Azure probe port resource - Azure DNS resource

Dynamic DNS resource

Parameters	Default value before update	Default value after update
Dynamic DNS resource Properties		
Dependency Tab		
• Follow the default dependence	On - floating IP resources - virtual IP resources - AWS elastic ip resource - AWS virtual ip resource - Azure probe port resource	On - floating IP resources - virtual IP resources - AWS elastic ip resource - AWS virtual ip resource - AWS secondary ip resource - Azure probe port resource

6.7.5 Moved Parameters

The following table shows the parameters which are configurable with Cluster WebUI but whose controls have been moved for each of the versions:

Internal Version 4.0.0-1

Before the change	After the change
[Cluster Properties]-[Recovery Tab]-[Max Reboot Count]	[Cluster Properties]-[Extension Tab]-[Max Reboot Count]
[Cluster Properties]-[Recovery Tab]-[Max Reboot Count Reset Time]	[Cluster Properties]-[Extension Tab]-[Max Reboot Count Reset Time]
[Cluster Properties]-[Recovery Tab]-[Use Forced Stop]	[Cluster Properties]-[Extension Tab]-[Use Forced Stop]

Continued on next page

Table 6.64 – continued from previous page

Before the change	After the change
[Cluster Properties]-[Recovery Tab]-[Forced Stop Action]	[Cluster Properties]-[Extension Tab]-[Forced Stop Action]
[Cluster Properties]-[Recovery Tab]-[Forced Stop Timeout]	[Cluster Properties]-[Extension Tab]-[Forced Stop Timeout]
[Cluster Properties]-[Recovery Tab]-[Virtual Machine Forced Stop Setting]	[Cluster Properties]-[Extension Tab]-[Virtual Machine Forced Stop Setting]
[Cluster Properties]-[Recovery Tab]-[Execute Script for Forced Stop]	[Cluster Properties]-[Extension Tab]-[Execute Script for Forced Stop]
[Cluster Properties]-[Auto Recovery Tab]-[Auto Return]	[Cluster Properties]-[Extension Tab]-[Auto Return]
[Cluster Properties]-[Exclusion Tab]-[Mount/Umount Exclusion]	[Cluster Properties]-[Extension Tab]-[Exclude Mount/Unmount Commands]
[Cluster Properties]-[Recovery Tab]-[Disable Recovery Action Caused by Monitor Resource Error]	[Cluster Properties]-[Extension Tab]-[Disable cluster operation]-[Recovery Action when Monitor Resource Failure Detected]
[Group Properties]-[Attribute Tab]-[Failover Exclusive Attribute]	[Group Common Properties]-[Exclusion Tab]

Internal Version 5.0.0-1

Before the change	After the change
[Cluster Properties]-[Extension Tab]-[Use Forced Stop]	[Cluster Properties]-[Fencing Tab]-[Forced Stop] - [Type]
[Cluster Properties]-[Extension Tab]-[Forced Stop Action]	[BMC Forced Stop Properties]-[Forced Stop Tab]-[Forced Stop Action]
[Cluster Properties]-[Extension Tab]-[Forced Stop Timeout]	[BMC Forced Stop Properties]-[Forced Stop Tab]-[Forced Stop Timeout]
[Cluster Properties]-[Extension Tab]-[Virtual Machine Forced Stop Setting] - [Action]	[vCenter Forced Stop Properties]-[Forced Stop Tab]-[Forced Stop Action]
[Cluster Properties]-[Extension Tab]-[Virtual Machine Forced Stop Setting] - [Timeout]	[vCenter Forced Stop Properties]-[Forced Stop Tab]-[Forced Stop Timeout]
[Cluster Properties]-[Extension Tab]-[Virtual Machine Forced Stop Setting] - [Host Name]	[vCenter Forced Stop Properties]-[vCenter Tab]-[Host Name]
[Cluster Properties]-[Extension Tab]-[Virtual Machine Forced Stop Setting] - [User Name]	[vCenter Forced Stop Properties]-[vCenter Tab]-[User Name]
[Cluster Properties]-[Extension Tab]-[Virtual Machine Forced Stop Setting] - [Password]	[vCenter Forced Stop Properties]-[vCenter Tab]-[Password]
[Cluster Properties]-[Extension Tab]-[Virtual Machine Forced Stop Setting] - [Perl Path]	[vCenter Forced Stop Properties]-[vCenter Tab]-[Perl Path]
[Server Properties]-[BMC Tab]-[IP Address]	[BMC Forced Stop Properties]-[Server List Tab]-[BMC Settings]-[IP Address]
[Server Properties]-[BMC Tab]-[User Name]	[BMC Forced Stop Properties]-[Server List Tab]-[BMC Settings]-[User Name]
[Server Properties]-[BMC Tab]-[Password]	[BMC Forced Stop Properties]-[Server List Tab]-[BMC Settings]-[Password]

CHAPTER

SEVEN

UPGRADING EXPRESSCLUSTER

This chapter provides information on how to upgrade EXPRESSCLUSTER.

This chapter covers:

7.1 How to upgrade from EXPRESSCLUSTER

7.1.1 How to upgrade from X3.3 or X4.x to X 5.0

Before starting the upgrade, read the following notes.

- The upgrade procedure described in this section is valid for EXPRESSCLUSTER X 3.3 for Linux (internal version 3.3.5-1) or later.
- In EXPRESSCLUSTER X 4.2 for Linux or later, port numbers for EXPRESSCLUSTER have been added. If you upgrade from EXPRESSCLUSTER X 4.1 for Linux or earlier, make necessary ports accessible beforehand.

  For information on port numbers for EXPRESSCLUSTER, refer to "6.3.1. *Communication port number*".
- If mirror disk resources or hybrid disk resources are set, cluster partitions require space of 1024 MiB or larger. And also, executing full copy of mirror disk resources or hybrid disk resources is required.
- If mirror disk resources or hybrid disk resources are set, it is recommended to backup data in advance. For details of a backup procedure, refer to "Backup procedures" and "Restoration procedures" in "Verifying operation" in the "Installation and Configuration Guide".
- Upgrade the EXPRESSCLUSTER Server RPM as root user.

See also:

For the procedure of updating between the different versions of the same major version, refer to the "Update Procedure Manual".

The following procedures explain how to upgrade from EXPRESSCLUSTER X 3.3 or 4.x to EXPRESSCLUSTER X 5.0.

1. Before upgrading, confirm that the servers in the cluster and all the resources are in normal status by using Cluster WebUI, WebManager or the command.
2. Save the current cluster configuration file with the Cluster WebUI, Builder or clpcfctrl command. For details about saving the cluster configuration file with clpcfctrl command, refer to "Changing, backing up, and checking cluster configuration data (clpcfctrl command)" -> "Backing up the cluster configuration data" in "EXPRESSCLUSTER command reference" in the "Reference Guide".
3. Uninstall the EXPRESSCLUSTER Server from all the servers. For details about uninstallation, refer to "Uninstallation" in "Uninstalling and reinstalling EXPRESSCLUSTER" in the "Installation and Configuration Guide".
4. Install the EXPRESSCLUSTER Server on all the servers. For details, refer to "Setting up the EXPRESSCLUSTER Server" in "Installing EXPRESSCLUSTER" and "Registering the license" in the "Installation and Configuration Guide".
5. In any of the servers with EXPRESSCLUSTER installed as above, execute the command for converting cluster configuration data.
   a. Move to the work directory (such as /tmp) in which the conversion command is to be executed.
   b. To the moved work directory, copy and deploy the cluster configuration data backed up in step 2.
      Deploy clp.conf and the scripts directory.

      Note:

      If backed up on Cluster WebUI, the cluster configuration data is zipped.
      Unzip the file, and clp.conf and the scripts directory will be extracted.

c. Execute the following command to convert the cluster configuration data:

```
# clpcfconv.sh -i .
```

d. Under the work directory, zip the cluster configuration data (clp.conf) and the scripts directory.

Note: Create the zip file so that when unzipped, the clp.conf file and scripts directory are created.

---

6. Open the config mode of Cluster WebUI, and click **Import**.
   Import the cluster configuration data zipped in step 5.
7. Of the cluster configuration data, manually update its items if necessary.
   See "6.7.2. *Removed Functions*". Then, if you have used any of the functions with its corresponding action described in the Action column of the table, change the cluster configuration data according to the described action.
8. If you upgrade from EXPRESSCLUSTER X 3.3 and are using mirror disk resources or hybrid disk resources, perform the following steps:
   a. Allocate cluster partition (The cluster partition should be 1024 MiB or larger).
   b. If the cluster partition is different from the configuration, modify the configuration. And regarding the groups which mirror disk resources or hybrid disk resources belong to, if Startup Attribute is Auto Startup on the Attribute tab of Group Properties, change it to Manual Startup.
   c. If mirror disk resources are set, perform the following steps for each mirror disk resource.
      - Click **Tuning** on the **Details** tab of **Resouce Properties**. Then, **Mirror disk resource tuning properties** dialog box is displayed.
      - Uncheck **Execute the initial mirror construction** on **Mirror** tab of the **Mirror disk resource tuning properties** dialog box.
9. If the forced stop function or the forced stop script is used, perform the following steps:
   a. Set the **Type** of **Forced Stop** on the **Fencing** tab of **Cluster Properties**.
      If the forced stop script is used: Select **Custom**.
      If the forced stop script is not used: With EXPRESSCLUSTER operated on physical machines, select **BMC**; with EXPRESSCLUSTER operated on virtual machines, select **vCenter**.
   b. Click **Properties** to display the properties window for the forced stop resource, and set each parameter.
10. Click **Export** of the Cluster WebUI to apply the configuration data.
    If using a fixed-term license, execute the following command:

```
clplcnsc --distribute
```

11. If you upgrade from EXPRESSCLUSTER X 3.3 and are using mirror disk resources or hybrid disk resources, perform the following steps:

    Initialize the cluster partition of all mirror disk resources and hybrid disk resources as below on each server.
    For the mirror disk

    ```
    clpmdinit --create force <mirror disk resource name>
    ```

    For the hybrid disk

    ```
    clphdinit --create force <hybrid disk resource name>
    ```

12. Open the operation mode of Cluster WebUI, and start the cluster.
13. If you upgrade from EXPRESSCLUSTER X 3.3 and are using mirror disk resources or hybrid disk resources, perform the following steps:
    a. Execute a full copy assuming that the server with the latest data is the copy source from the Mirror disks.
    b. Start the groups and confirm that each resource starts normally.
    c. If **Startup Attribute** or **Execute the initial mirror construction** was changed in the step 8, change back the setting with Cluster WebUI andclick **Export** to apply the cluster configuration data to the cluster.
14. This completes the procedure for upgrading the EXPRESSCLUSTER Server. Check that the servers are operating normally as the cluster by the clpstat command or Cluster WebUI

CHAPTER

EIGHT

GLOSSARY

Cluster partition

A partition on a mirror disk. Used for managing mirror disks.
(Related term: Disk heartbeat partition)

Interconnect

A dedicated communication path for server-to-server communication in a cluster.
(Related terms: Private LAN, Public LAN)

Virtual IP address IP address used to configure a remote cluster.

Management client Any machine that uses the Cluster WebUI to access and manage a cluster system.

Startup attribute A failover group attribute that determines whether a failover group should be started up automatically or manually when a cluster is started.

Shared disk A disk that multiple servers can access.

Shared disk type cluster A cluster system that uses one or more shared disks.

Switchable partition

A disk partition connected to multiple computers and is switchable among computers.
(Related terms: Disk heartbeat partition)

Cluster system Multiple computers are connected via a LAN (or other network) and behave as if it were a single system.

Cluster shutdown To shut down an entire cluster system (all servers that configure a cluster system).

Active server

A server that is running for an application set.
(Related term: Standby server)

Secondary server

A destination server where a failover group fails over to during normal operations.
(Related term: Primary server)

Standby server

A server that is not an active server.
(Related term: Active server)

Disk heartbeat partition A partition used for heartbeat communication in a shared disk type cluster.

Data partition

A local disk that can be used as a shared disk for switchable partition. Data partition for mirror disks and hybrid disks.

(Related term: Cluster partition)

Network partition

All heartbeat is lost and the network between servers is partitioned.

(Related terms: Interconnect, Heartbeat)

Node A server that is part of a cluster in a cluster system. In networking terminology, it refers to devices, including computers and routers, that can transmit, receive, or process signals.

Heartbeat

Signals that servers in a cluster send to each other to detect a failure in a cluster.

(Related terms: Interconnect, Network partition)

Public LAN

A communication channel between clients and servers.

(Related terms: Interconnect, Private LAN)

Failover The process of a standby server taking over the group of resources that the active server previously was handling due to error detection.

Failback A process of returning an application back to an active server after an application fails over to another server.

Failover group A group of cluster resources and attributes required to execute an application.

Moving failover group Moving an application from an active server to a standby server by a user.

Failover policy A priority list of servers that a group can fail over to.

Private LAN

LAN in which only servers configured in a clustered system are connected.

(Related terms: Interconnect, Public LAN)

Primary (server)

A server that is the main server for a failover group.

(Related term: Secondary server)

Floating IP address

Clients can transparently switch one server from another when a failover occurs.

Any unassigned IP address that has the same network address that a cluster server belongs to can be used as a floating address.

Master server The server displayed at the top of Master Server in Server Common Properties of the Cluster WebUI

Mirror disk connect LAN used for data mirroring in mirror disks and hybrid disks. Mirror connect can be used with primary interconnect.

Mirror disk type cluster A cluster system that does not use a shared disk. Local disks of the servers are mirrored.

CHAPTER

NINE

INDEX

CHAPTER

TEN

LEGAL NOTICE

10.1 Disclaimer

- Information in this document is subject to change without notice. No part of this document may be reproduced or transmitted in any form by any means, electronic or mechanical, for any purpose, without the express written permission of NEC Corporation.

10.2 Trademark Information

- EXPRESSCLUSTER® is a registered trademark of NEC Corporation.
- Linux is a registered trademark of Linus Torvalds in the United States and other countries.
- Microsoft, Windows, Windows Server, Internet Explorer, Azure, and Hyper-V are registered trademarks of Microsoft Corporation in the United States and other countries.
- Firefox is a trademark or registered trademark of Mozilla Foundation.
- Google Chrome is a trademark or registered trademark of Google, Inc.
- Google Cloud Platform (GCP) is a trademark or a registered trademark of Google LLC.
- SUSE is a registered trademark of SUSE LLC in the United States and other countries.
- Asianux is registered trademark of Cybertrust Japan Co., Ltd. in Japan
- Ubuntu is a registered trademark of Canonical Ltd.
- Amazon Web Services and all AWS-related trademarks, as well as other AWS graphics, logos, page headers, button icons, scripts, and service names are trademarks, registered trademarks or trade dress of AWS in the United States and/or other countries.
- Apache Tomcat, Tomcat, and Apache are registered trademarks or trademarks of Apache Software Foundation.
- Citrix, Citrix XenServer, and Citrix Essentials are registered trademarks or trademarks of Citrix Systems, Inc. in the United States and other countries.
- VMware, vCenter Server, and vSphere is registered trademarks or trademarks of VMware, Inc. in the United States and/or other jurisdictions.
- Python is a registered trademark of the Python Software Foundation.
- SVF is a registered trademark of WingArc Technologies, Inc.
- JBoss is a registered trademark of Red Hat, Inc. or its subsidiaries in the United States and other countries.
- Oracle, Oracle Database, Solaris, MySQL, Tuxedo, WebLogic Server, Container, Java, and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle Corporation and/or its affiliates.
- SAP, SAP NetWeaver, and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries.
- IBM, DB2, and WebSphere are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both.
- MariaDB is a registered trademark of MariaDB Corporation AB.
- PostgreSQL is a registered trademark of the PostgreSQL Global Development Group.
- PowerGres is a registered trademark of SRA OSS, Inc.
- Sybase is a registered trademark of Sybase, Inc.
- RPM is a registered trademark of Red Hat, Inc. or its subsidiaries in the United States and other countries.
- F5, F5 Networks, BIG-IP, and iControl are trademarks or registered trademarks of F5 Networks, Inc. in the United States and other countries.
- Equalizer is a registered trademark of Coyote Point Systems, Inc.
- WebOTX is a registered trademark of NEC Corporation.
- WebSAM is a registered trademark of NEC Corporation.

- Other product names and slogans written in this manual are trademarks or registered trademarks of their respective companies.

CHAPTER

ELEVEN

REVISION HISTORY

Edition	Revised Date	Description
1st	Apr 08, 2022	New manual
2nd	Apr 26, 2022	Corresponds to the internal version 5.0.1-1.
3rd	Jul 29, 2022	Corrected typographical errors.
4th	Nov 04, 2022	Corresponds to the internal version 5.0.2-1.

© Copyright NEC Corporation 2022. All rights reserved.