Copyright © 2012, Hitachi, Ltd.

THEORY OF OPERATION SECTION

Contents

Copyright © 2012, 2013, Hitachi, Ltd.

Copyright © 2012, Hitachi, Ltd.

Copyright © 2012, Hitachi, Ltd.

 Copyright © 2012, Hitachi, Ltd.

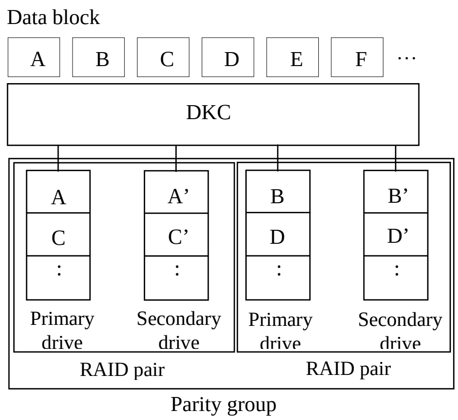
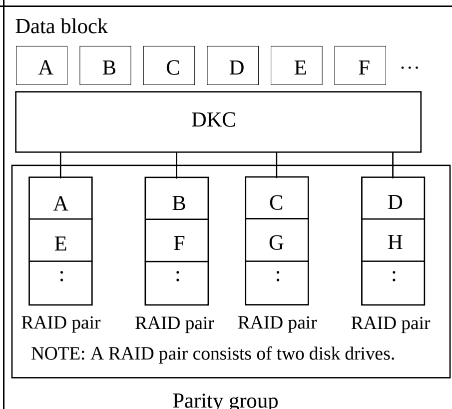
1. RAID Architecture Overview

The objectives of the RAID technology are the low cost, high reliability, and high I/O performance of disk storage devices. To achieve these objectives, this storage system supports levels 1, 5 and 6 of RAID technologies (in this section, part of level 3 RAID technology is explained to make the outline of RAID5 more understandable). The features of the levels of RAID technologies are described below.

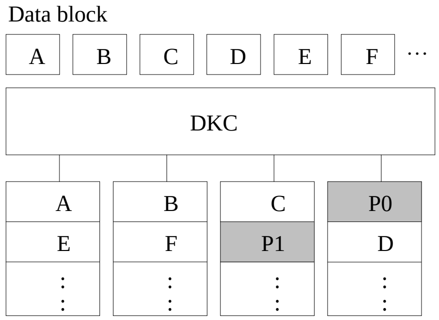
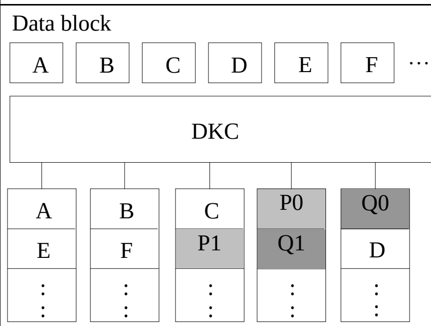
1.1 Outline of RAID Systems

The concept of disk array was announced in 1987 by the research group of University of California at Berkeley.
The research group called the disk array RAID (Redundant Array of Inexpensive Disks: A storage system that has redundancy by employing multiple inexpensive and small disk drives), classified the RAID systems into five levels, that is, RAID 1 to RAID 5, and added RAID 0 and RAID 6 later.
Since the DKC710I storage system supports RAID 1, RAID 5, and RAID 6, the method, advantage, and disadvantage of each of them are explained below.

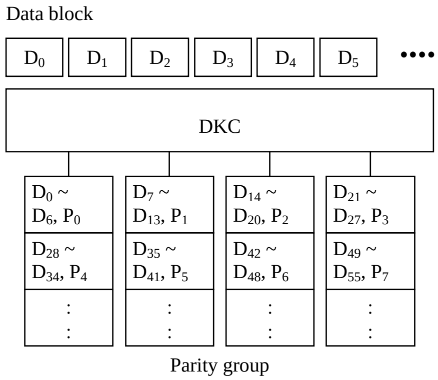
Table 1.1-1 Outline of RAID Systems

Level	Configuration		Characteristics
RAID 1 (2D+2D) configuration		Outline	Mirror disks (duplicated writing) Two disk drives, primary and secondary disk drives, compose a RAID pair (mirroring pair) and the identical data is written to the primary and secondary disk drives. Further, data is scattered on the two RAID pairs.
		Advantage	RAID 1 is highly usable and reliable because of the duplicated data. It has higher performance than ordinary RAID 1 (when it consists of two disk drives) because it consists of the two RAID pairs.
		Disadvantage	A disk capacity twice as large as user data capacity is required.
RAID 1 (4D+4D) configuration (Two concatenation of RAID1 (2D+2D))		Outline	Mirror disks (duplicated writing) The two parity groups of RAID 1 (2D+2D) are concatenated and data is scattered on them. In the each RAID pair, data is written in duplicate.
		Advantage	This configuration is highly usable and reliable because of the duplicated data. It has higher performance than the 2D+2D configuration because it consists of the four RAID pairs.
		Disadvantage	A disk capacity twice as large as user data capacity is required.

 Copyright © 2012, Hitachi, Ltd.

RAID5	 Data disks + Parity disk NOTE: There are two configurations of RAID 5: 3D+1P configuration (four disk drives) and 7D+1P configuration (eight disk drives). The above diagram shows the 3D+1P configuration. In the 7D+1P configuration, data is arranged in the same way.	Outline	Data is written to multiple disks successively in units of block (or blocks). Parity data is generated from data of multiple blocks and written to optional disk.
		Advantage	RAID 5 fits the transaction operation mainly uses small size random access because each disk can receive I/O instructions independently. It can provide high reliability and usability at a comparatively low cost by virtue of the parity data.
		Disadvantage	Write penalty of RAID 5 is larger than that of RAID 1 because pre-update data and pre-update parity data must be read internally because the parity data is updated when data is updated.
RAID6	 Data disks + Parity disks P and Q NOTE: RAID 6 (6D+2P) configuration practically consists of eight disk drives. RAID 6 (14D+2P) configuration practically consists of sixteen disk drives. In the above diagram, three disk drives are omitted.	Outline	Data blocks are scattered to multiple disks in the same way as RAID 5 and two parity disks, P and Q, are set in each row. Therefore, data can be assured even when failures occur in up to two disk drives in a parity group.
		Advantage	RAID 6 is far more reliable than RAID 1 and RAID 5 because it can restore data even when failures occur in up to two disks in a parity group.
		Disadvantage	Because the parity data P and Q must be updated when data is updated, RAID 6 is imposed write penalty heavier than that on RAID 5, performance of the random writing is lower than that of RAID 5 in the case where the number of drives makes a bottleneck.

Copyright © 2012, Hitachi, Ltd.

<table>
<tr><td rowspan="3">RAID5 concatenation</td><td rowspan="3">

NOTE:
The above-mentioned figure is four concatenation cofiguration, but it is the same in the case of two concatenation.</td><td>Outline</td><td>In the case of RAID5 (7D+1P), two or four parity groups (eight drives) are concatenated, and the data is distributed and arranged in 16 drives or 32 drives.</td></tr>
<tr><td>Advantage</td><td>When the parity group becomes a performance bottleneck, the performance improvement can be attempted because it is configured with twice and four times the number of drives in comparison with RAID5 (7D+1P).</td></tr>
<tr><td>Disadvantage</td><td>The influence level when two drives are blocked is large because twice and four times LDEVs are arranged in comparison with RAID5 (7D+1P). However, the probability that the read of the single block in the parity group becomes impossible due to the failure is the same as that of RAID5 (7D+1P).</td></tr>
</table>

Copyright © 2012, Hitachi, Ltd.

1.2 Comparison of RAID levels

(1) Space efficiency

RAID level	Space efficiency (User area/Disk capacity)	Remarks
RAID1 2D+2D	50.0%	Because of the mirroring
RAID1 4D+4D	50.0%	Because of the mirroring
RAID5 3D+1P	75.0%	Ratio of the number of parity disks to the number of data disks The space efficiency of the 6D+2P is the same as that of the 3D+1P. Two concatenation and four concatenation of 7D+1P are also the same.
RAID5 7D+1P	87.5%	
RAID6 6D+2P	75.0%	
RAID6 14D+2P	87.5%	

(2) Comparison of performance limits of parity groups (When supposing the marginal efficiency of the RAID 1 (2D+2D) to be 100%)

RAID level	Random and sequential reading	Sequential writing	Random writing
RAID1 2D+2D	100%	100%	100%
RAID1 4D+4D	200%	200%	200%
RAID5 3D+1P	100%	150%	50%
RAID5 7D+1P	200%	350%	100%
RAID6 6D+2P	200%	300%	66.7% (The efficiency is lowered by 33% compared with the 7D+1P.)
RAID6 14D+2P	400%	700%	133.4%
Remarks	Proportionate to the number of HDDs	Proportionate to the number of data HDDs	See the explanation below.

Copyright © 2012, Hitachi, Ltd.

- In the case of two concatenation and four concatenation RAID5 (7D+1P), it becomes the value twice and four times the above-mentioned.
- The reason why the efficiency is lowered by 33% in the case of RAID 6 (6D+2P) in comparison with RAID 5 (7D+1P) is as follows.

When RAID 5 executes random writing, it issues a total of four IOs, that is, reading of old data, reading of old parity data, writing of new data, and writing of new parity data to disk drives.
In the case of RAID 6, on the other hand, it issues a total of six IOs, that is, reading of old data, reading of old parity data (P), reading of old parity data (Q), writing of new data, writing of new parity data (P), and writing of new parity data (Q) to disk drives.
The number of IOs that RAID 5 issues is four, whereas those that RAID 6 issues is six; the latter is 1.5 times as many as the former. Therefore, the random writing performance of RAID 6 is lowered by 33% in comparison with RAID 5.

However, unless RAID 6 is in an environment in which the number of drives makes a bottleneck, the write penalty is absorbed by the cache memory, so that the performance is not lowered.

 Copyright © 2012, Hitachi, Ltd.

(3) Reliability

<table>
<tr><th>RAID level</th><th>Conditions of data assurance</th></tr>
<tr><td>RAID1
2D+2D</td><td rowspan="2">When a failure occurs in one of the mirroring pair of disk drives, data can be restored through use of data of the other disk drives.
When failures occur in both of the mirroring pair of disk drives, an LDEV blockade is caused.</td></tr>
<tr><td>RAID1
4D+4D</td></tr>
<tr><td>RAID5
3D+1P</td><td rowspan="2">When a failure occurs in one disk drive in a parity group, data can be restored through use of the parity data.
When failures occur in two disk drives, an LDEV blockade is caused.</td></tr>
<tr><td>RAID5
7D+1P</td></tr>
<tr><td>RAID6
6D+2P</td><td rowspan="2">When the failure(s) occur(s) in one or two disk drive(s) in a parity group, data can be restored through use of the parity data. When three disk drives fail, an LDEV blockade is caused.</td></tr>
<tr><td>RAID6
14D+2P</td></tr>
</table>

In the case of RAID 6, data can be assured when up to two drives in a parity group fail, as explained above. Therefore, RAID 6 is the most reliable in the RAID levels.

2. Hardware Specifications

2.1 General

DW700 is a new storage system of a high midrange-class that offers high performance of the enterprise-class.
DW700 consists of the Controller Chassis and the Drive Box as well as the storage of the midrange-class, and they are installed in a 19-inch rack. All drives are installed in the Drive Box without installing the drive in the Controller Chassis, and up to 1,152 drives are supported.
Dual controller architecture is adopted in the controller part that is installed in the Controller Chassis. The channel I/F supports only the open system, and the mainframe is not supported. The power supply is single phase AC 200V.

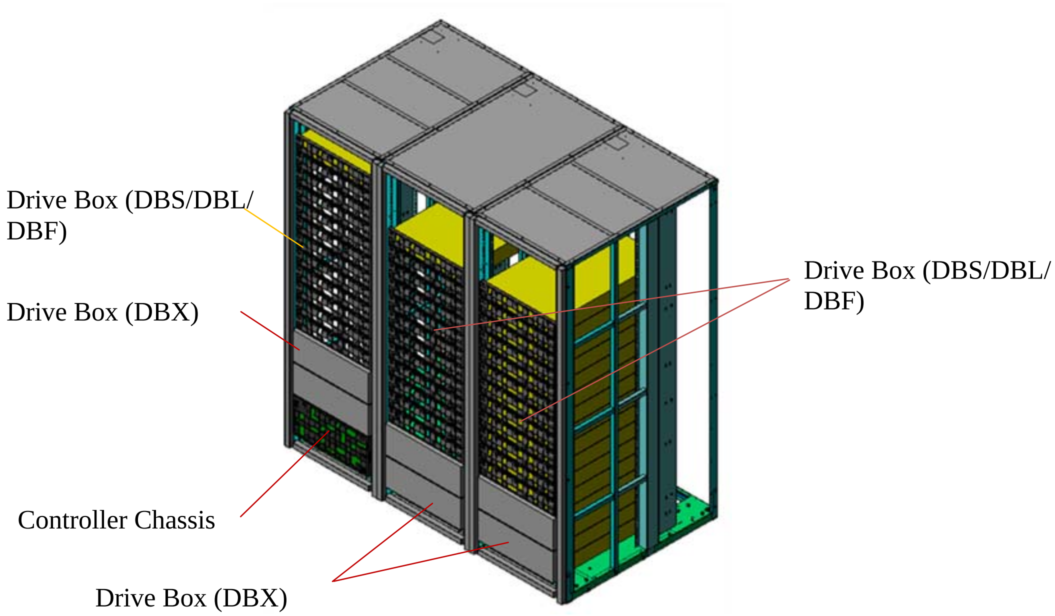


Fig. 2.1-1 DW700 Storage System

Copyright © 2012, 2013, Hitachi, Ltd.

2.1.1 Features

(1) Scalability

DW700 provides variations of the storage system configuration according to the kinds and the numbers of selected options: channel blades, cache memory, disk drive, flash drive and flash module drive (FMD).

- Number of installed channel options: 2 to 8 (Up to 12 when the DKB slot is used)
- Capacity of cache memory: 32GB to 256GB
- Number of disk drives: Up to 1,152 (2.5-inch and 3.5-inch HDDs) or 96 (FMD)

(2) High-performance

- DW700 supports three kinds of high-speed disk drives at the speed of 7,200 min^{-1}, 10,000 min^{-1} and 15,000 min^{-1}.
- DW700 supports flash drives with ultra high-speed response.
- DW700 supports Flash Module Drive (FMD) with ultra high-speed response and high capacity.
- The high-speed data transfer between DKB and HDDs at a rate of 6Gbps with the SAS interface is achieved.
- DW700 uses Intel processor with brand new technology which performs as excellent as that of the enterprise device DKC710I.

(3) Large Capacity

- DW700 supports disk drives with capacities of 300GB, 600GB, 900GB, 1.2TB, 3TB and 4TB.
- DW700 supports flash drives with capacities of 200GB, 400GB and 800GB.
- DW700 supports Flash Module Drive (FMD) with capacities of 1.6TB and 3.2TB.
- DW700 controls up to 16,384 logical volumes and up to 1,152 disk drives, and provides a physical disk capacity of approximately 4,511TB per storage system.

(4) Flash Module Drive (FMD)

The FMD is a flash drive with large capacity which has been accomplished by adopting the Hitachi original package.

Its interface is 6Gbps SAS same as that of the HDD/SDD.

The FMD uses MLC-NAND Flash Memory and features high performance, long service life and superior cost performance by virtue of its original control methods.

Copyright © 2012, 2013, Hitachi, Ltd.

(5) Connectivity

DW700 supports OS's for various UNIX servers and PC servers, so that it conforms to heterogeneous system environment in which those various OS's coexist.
The platforms that can be connected are shown in the following table.

Table 2.1.1-1 Support OS Type

Maker	OS
HP	HP-UX
	Tru64
	OpenVMS
Sun	Solaris
IBM	AIX 5L
Microsoft	Windows
NOVELL	NetWare
	SUSE Linux
Red Hat	Red Hat Linux
VMware	ESX Server

A channel interface supported by the DW700 is shown below.
- fibre channel

(6) High reliability

- DW700 supports RAID6 (6D+2P/14D+2P), RAID5 (3D+1P/7D+1P), and RAID1 (2D+2D/4D+4D).
- Main components are implemented with a duplex or redundant configuration, so even when single point of the component failure has occurred, the storage system can continue the operation.
  However, when the failure of the main board with the cache memory is maintained, the channel ports and the drive ports of the cluster concerned are blocked.

(7) Non-disruptive Service and Upgrade

- Main components can be added, removed, and replaced without shutting down a device while the storage system is in operation.
  However, when the additional installation of the cache memory is executed, the channel ports and the drive ports of the cluster concerned are blocked.
- A Service Processor (SVP) mounted on Controller Chassis monitors the running condition of the storage system. Connecting the SVP with a service center enables remote maintenance.
- The microcode can be upgraded without shutting down the storage system.

 Copyright © 2012, 2013, Hitachi, Ltd.

2.2 Architecture

2.2.1 Outline

DW700 consists of the Controller Chassis installed with control boards and the Drive Box to be installed with various types of HDDs. The Controller Chassis and the Drive Box are mounted in a 19-inch rack.
There are four kinds of Drive Boxes. The Drive Box (DBS) can be installed with up to 24 2.5-inch HDDs. The Drive Box (DBL) can be installed with up to 12 3.5-inch HDDs. The Drive Box (DBX) can be installed with up to 24 3.5-inch HDDs. The Drive Box (DBF) can be installed with up to 12 FMDs.
A Controller Chassis controls up to 48 Drive Box (DBS/DBL). A Controller Chassis controls up to 24 Drive Box (DBX). A Controller Chassis controls up to 8 Drive Box (DBF). The Drive Box (DBS), the Drive Box (DBL), the Drive Box (DBX) and the Drive Box (DBF) can be mixed in the storage system. A Controller Chassis can control up to 1,152 HDDs.
The size of each unit is as follows: The Controller Chassis is 5U high, the Drive Box (DBS/DBL/DBF) is 2U high and the Drive Box (DBX) is 4U high.
The minimum configuration of the storage system consists of one Controller Chassis and one Drive Box because DW700 allows free allocation of HDDs to make a RAID group. However, to configure the storage system with superb performance, it is recommended to install or add the Drive Box (DBS/DBL/DBF) four pieces at a time or the Drive Box (DBX) two pieces at a time per Controller Chassis.

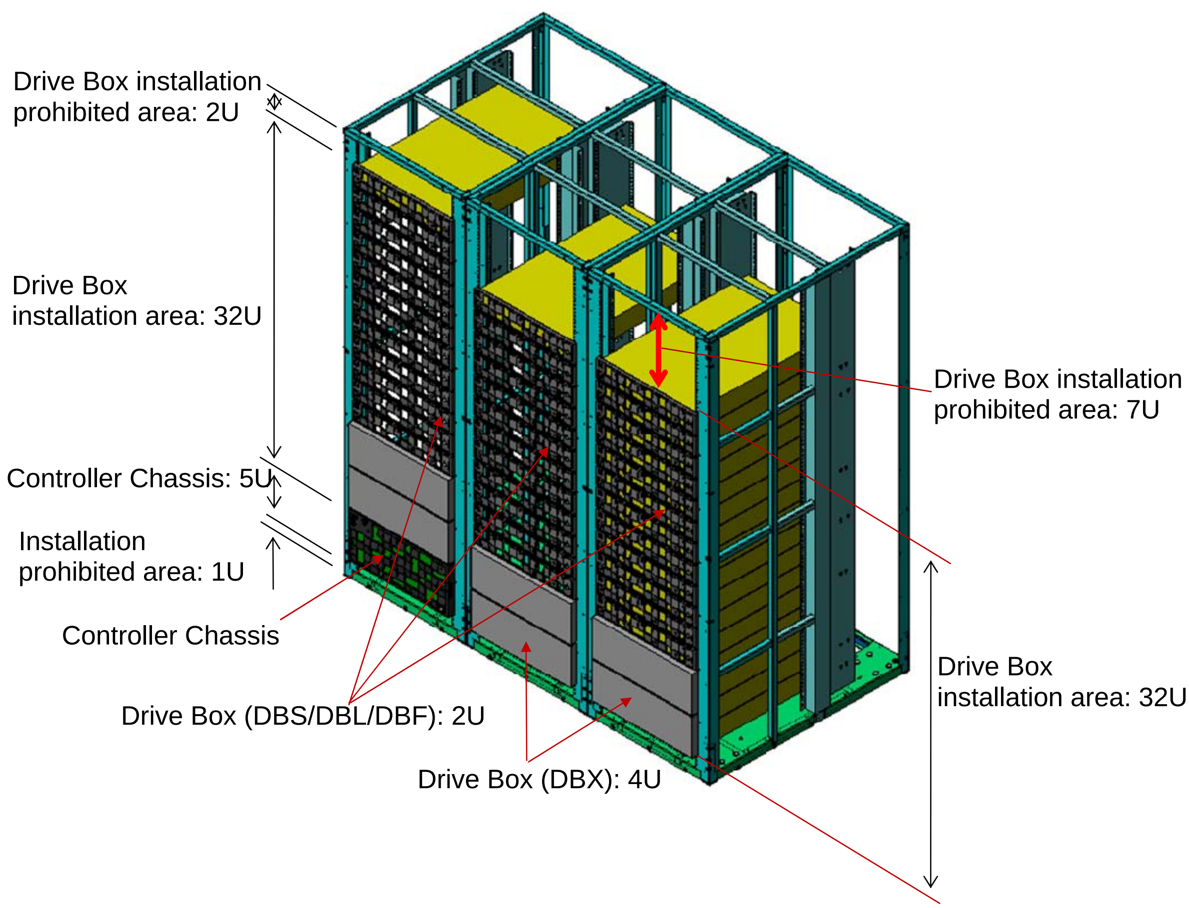


Fig. 2.2.1-1 DW700 Outline

 Copyright © 2012, 2013, Hitachi, Ltd.

2.2.2 Hardware Architecture

(1) Controller Chassis (DKC)
The Controller Chassis (DKC) consists of MAIN Blades (MAIN), MP Blades (MPB), Host I/O Module (CHB), Backend I/O Module (DKB), Service Processor (SVP) and AC-DC Power Supply (SWPS) with a built-in cooling fan.
The cache memory is installed in the MAIN Blades.
The Battery and the Cache Flash Memory are also installed in the MAIN Blades to prevent data loss by the occurrence of power outage or the like.
The storage system continues to operate when a single point of failure occurred, by adopting a duplexed configuration for each control board (MAIN, MPB, CHB, DKB) and the AC-DC power supply, and a redundant configuration for the AC-DC power supply and the cooling fan.
The addition and the replacement of the components and the upgrading of the microcode can be processed while the storage system is in operation. However, when executing the maintenance and replacement of the MAIN Blades, the CHB and the DKB of the cluster concerned are blocked.
The SVP implements a setting and a modification of the storage system configuration information, and observes the operational status. Connecting this SVP to Service Center enables the remote maintenance of the storage system.

(2) Drive Box (DBS)
The Drive Box (DBS) is a chassis to install the 2.5-inch disk drives and the 2.5-inch flash drives, and consists of ENC, the cooling fan and the AC-DC power supply. The duplex configuration is adopted in ENC and AC-DC power supply, and the redundant configuration is adopted in AC-DC power supply and the cooling fan. All the components can be replaced and added while the storage system is in operation.

(3) Drive Box (DBL)
The Drive Box (DBL) is a chassis to install the 3.5-inch disk drives and the 3.5-inch flash drives, and consists of ENC, the cooling fan and the AC-DC power supply. The duplex configuration is adopted in ENC and AC-DC power supply, and the redundant configuration is adopted in AC-DC power supply and the cooling fan. All the components can be replaced and added while the storage system is in operation.

(4) Drive Box (DBX)
The Drive Box (DBX) is a chassis to install the 3.5-inch disk drives, and consists of ENC, the cooling fan and the AC-DC power supply. The duplex configuration is adopted in ENC and AC-DC power supply, and the redundant configuration is adopted in AC-DC power supply and the cooling fan. All the components can be replaced and added while the storage system is in operation.

 Copyright © 2012, 2013, Hitachi, Ltd.

(5) Drive Box (DBF)

The Drive Box (DBF) is a chassis to install the 3.5-inch disk drives, and consists of ENC, the cooling fan and the AC-DC power supply. The duplex configuration is adopted in ENC and AC-DC power supply, and the redundant configuration is adopted in AC-DC power supply and the cooling fan. All the components can be replaced and added while the storage system is in operation.

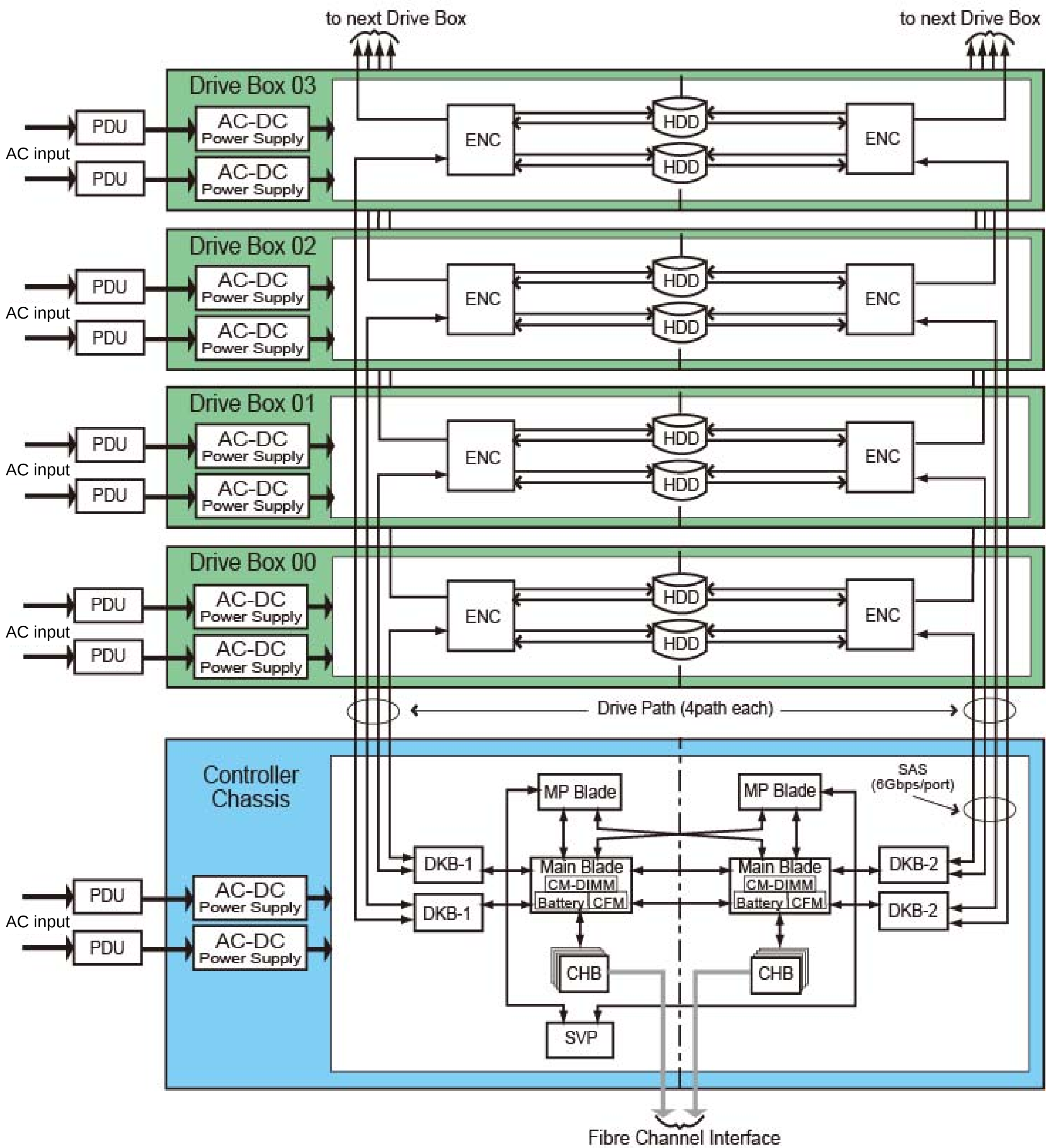


Fig. 2.2.2-1 DW700 Hardware Architecture Outlines

 Copyright © 2012, 2013, Hitachi, Ltd.

(6) Drive Path

(a) When using 2.5-inch HDD (SFF)
DW700 controls 1,152 HDDs with 8 paths.

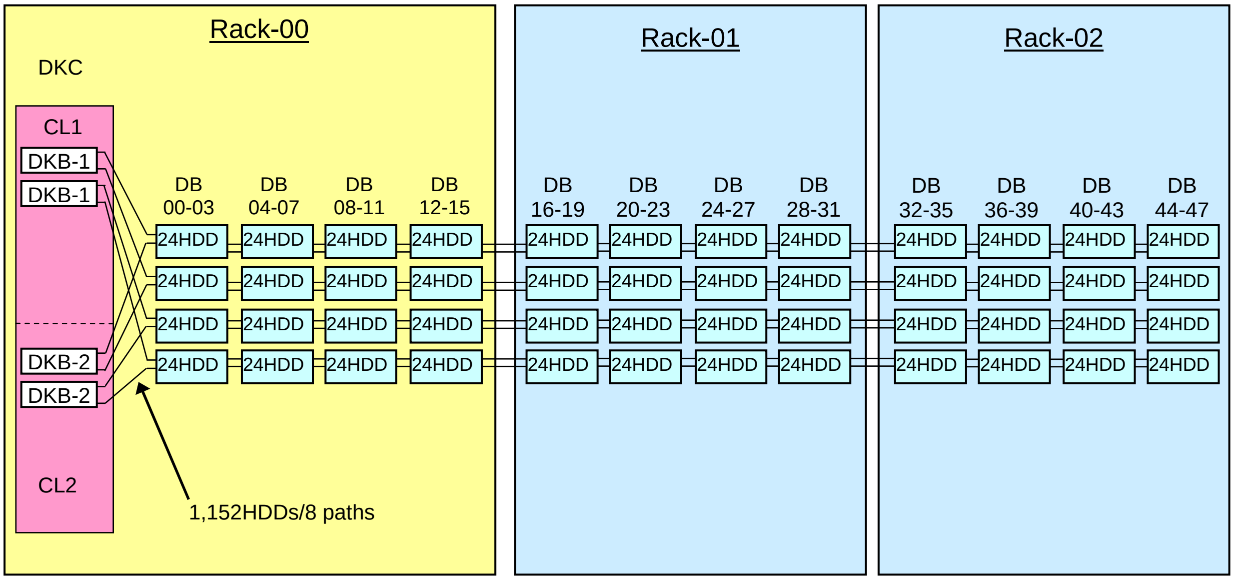


Fig. 2.2.2-2 Drive Path Connection Outlines when using 2.5-inch drives

(b) When using 3.5-inch HDD (LFF)
DW700 controls 576 HDDs with 8 paths.

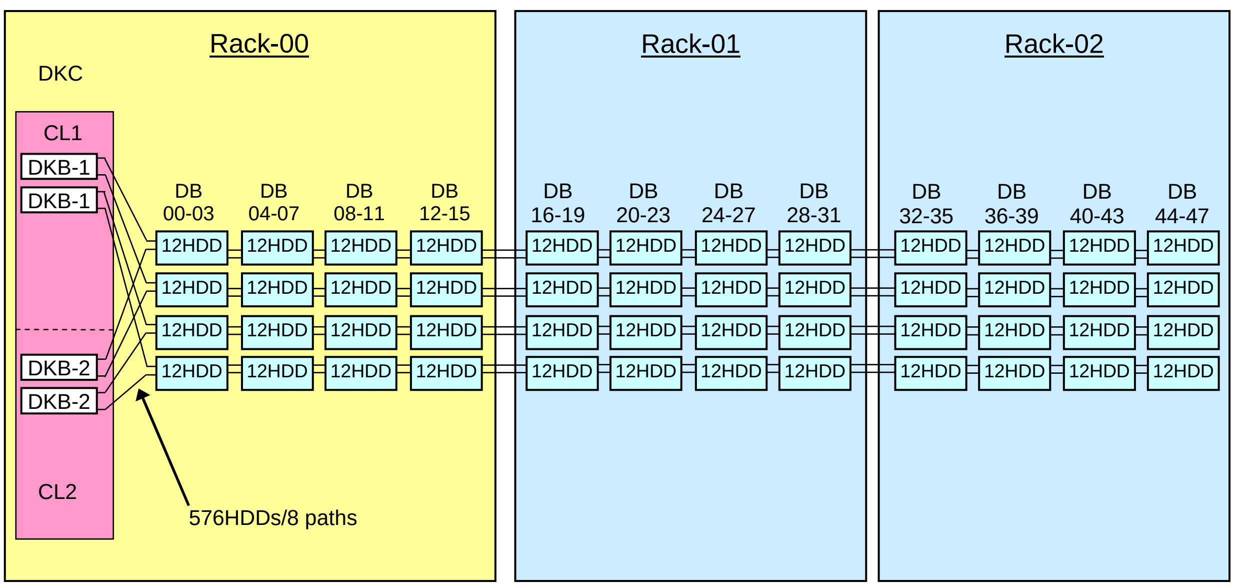


Fig. 2.2.2-3 Drive Path Connection Outlines when using 3.5-inch drives

 Copyright © 2012, 2013, Hitachi, Ltd.

(c) When using 3.5-inch HDD (DENSE)
DW700 controls 1,152 HDDs with 8 paths.

NOTICE:	Up to six DBX (DENSE) can be installed in a rack. Up to five DBX can be installed in a rack when a DKC (CBX) is installed there. Install the DBX at a height of 1,300mm or less above the ground (at a range between 2U and 26U).

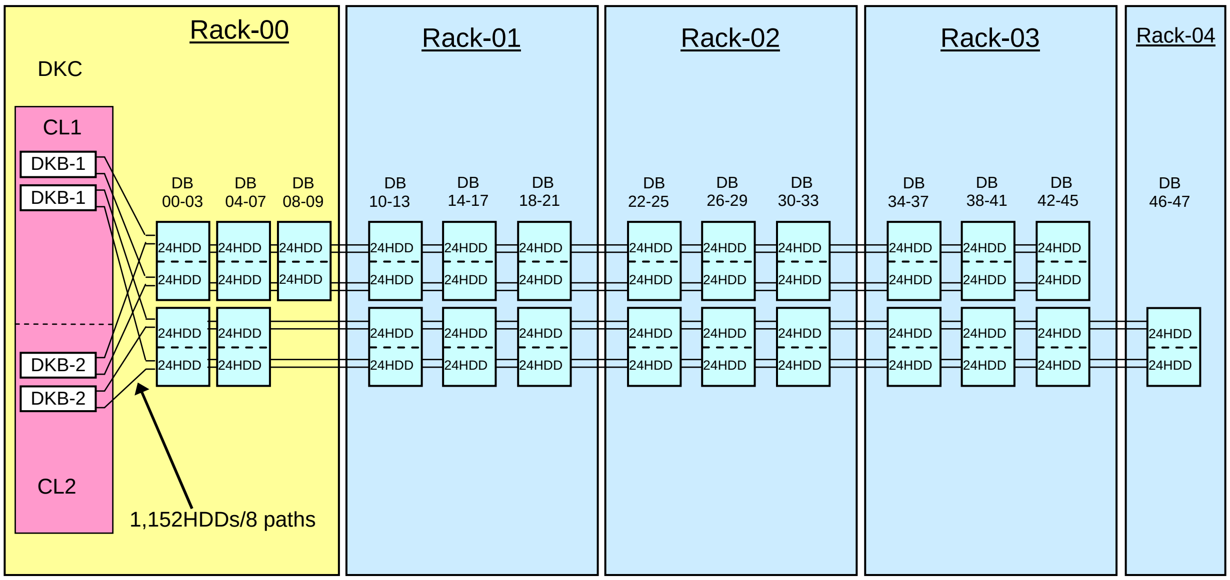


Fig. 2.2.2-4 Drive Path Connection Outlines when using 3.5-inch drives

(d) When using Flash Module Drive (FMD)
DW700 controls 96 FMDs with 8 paths.

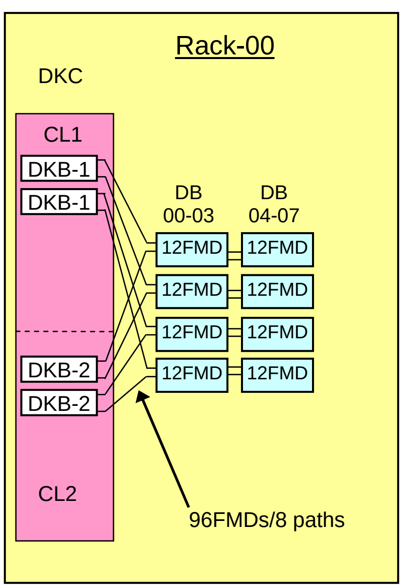


Fig. 2.2.2-5 Drive Path Connection Outlines when using FMDs

Copyright © 2012, Hitachi, Ltd.

2.2.3 Hardware Component

(1) Controller Chassis (DKC)

The Controller Chassis is standard installed with the Main Blades (MAIN), MP Blades (MPB), Service Processor (SVP) and AC-DC Power Supply (SWPS) with a built-in cooling fan. The required number of Backend I/O Module (DKB) and Host I/O Module (CHB) shall be installed according to system configuration.

If HDDs are used, Backend I/O Module (DKB) must be installed. For diskless configuration, Backend I/O Module (DKB) is not required. Host I/O Module (CHB) can be installed in four empty slots for Backend I/O Module (DKB).

Two or more Host I/O Module (CHB) must be installed. It is necessary to add Host I/O Module (CHB) two pieces at a time. Up to eight Host I/O Module (CHB) can be installed for configuration with disks.

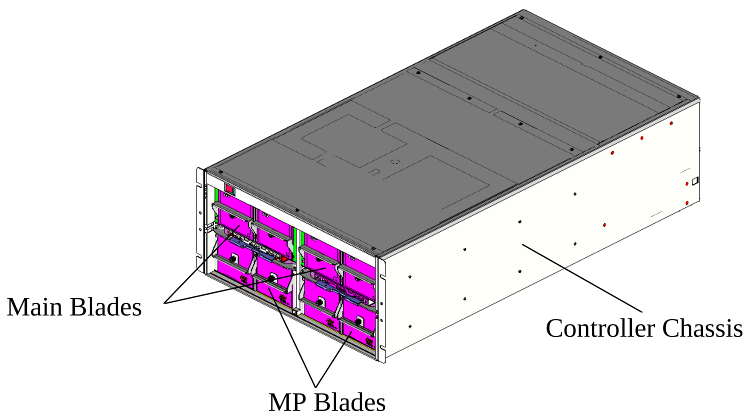


Fig. 2.2.3-1 Front View of Controller Chassis

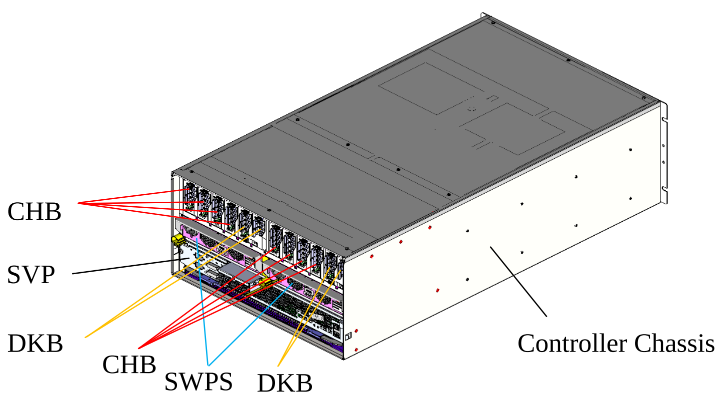


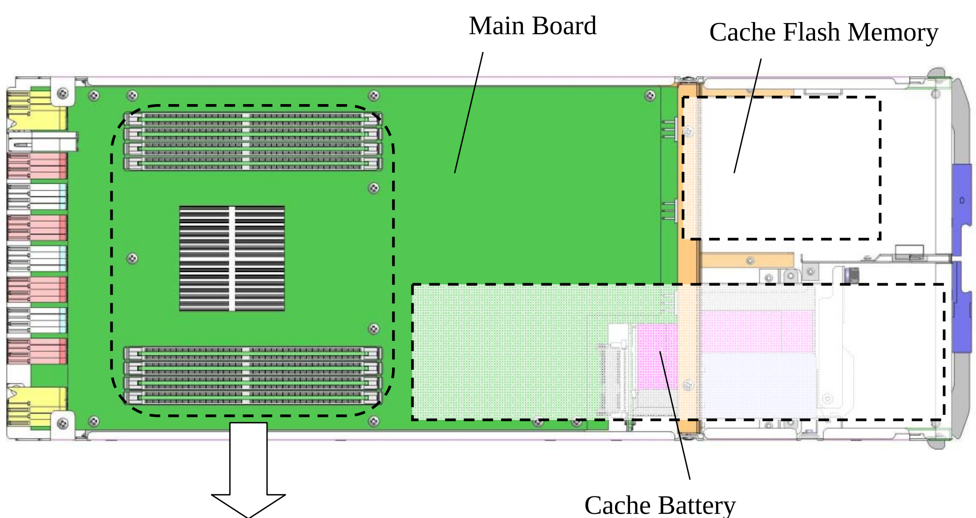
Fig. 2.2.3-2 Rear View of Controller Chassis

 Copyright © 2012, Hitachi, Ltd.

(2) MAIN Blades (MAIN)
The Cache Memory (CM-DIMM), the Cache Flash Memory, and the Battery for the Cache Flash Memory are installed in the Main Board.

Table 2.2.3-1 MAIN Blades Specifications

PCB Name	WP770-A
Number of PCB	1
Necessary number of PCB per controller chassis	2
Number of DIMM slot	8
Cache Memory Capacity	16 ~ 128GB
Number of Cache Flash Memory	1
Number of Cache Battery	1



CM DIMM Location

CM01:	*1
CM11:	*2
CM03:	*1
CM13:	*2

CM12:	*2
CM02:	*1
CM10:	*2
CM00:	*1

*1 belongs to CMG0 (Cache Memory Group 0).
*2 belongs to CMG1 (Cache Memory Group 1).
CMG0 must be always installed with CM-DIMM.
CMG1 is a slot to add CM-DIMM.
CM-DIMM must be installed in 4 piece set.
CM-DIMM of different capacity can be installed in CMG0 and CMG1.

Fig. 2.2.3-3 Top of MAIN Blade

 Copyright © 2012, Hitachi, Ltd.

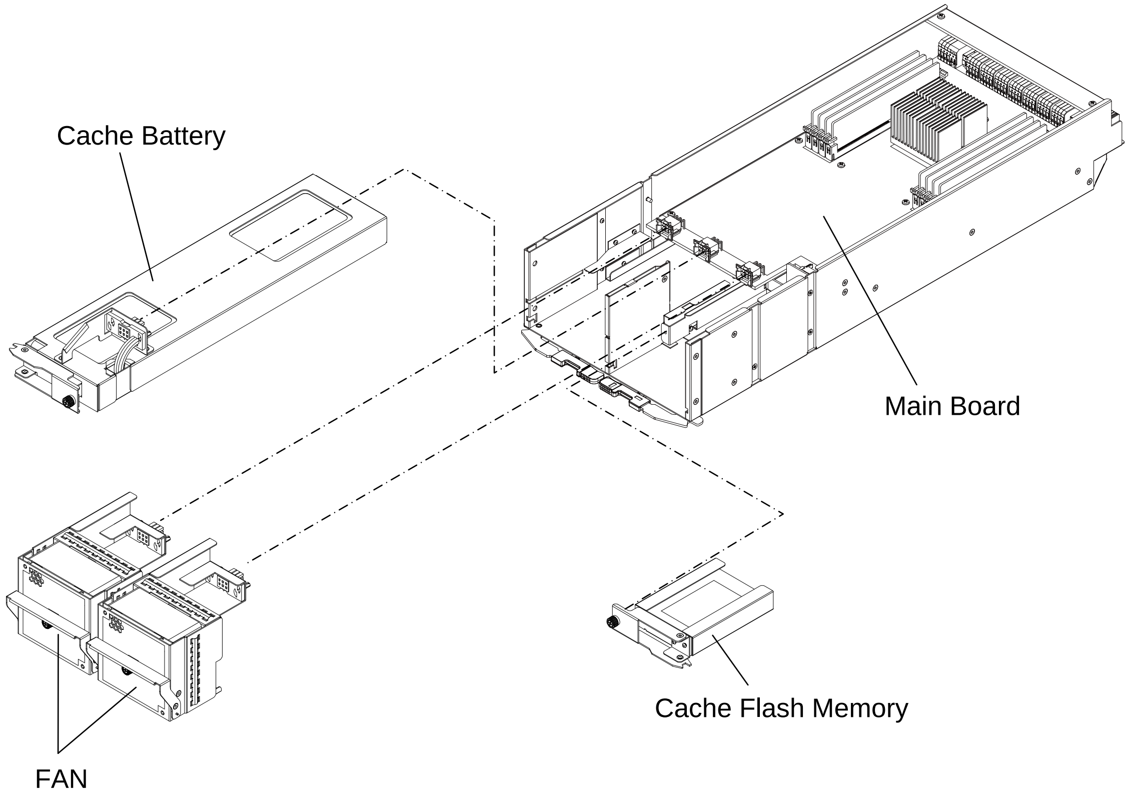


Fig. 2.2.3-4 MAIN Blade

 Copyright © 2012, 2013, Hitachi, Ltd.

(3) Cache Memory Module (CM-DIMM)
DW700 can be used the three kinds of CM-DIMM capacity.

Table 2.2.3-2 Cache Memory Module Specifications

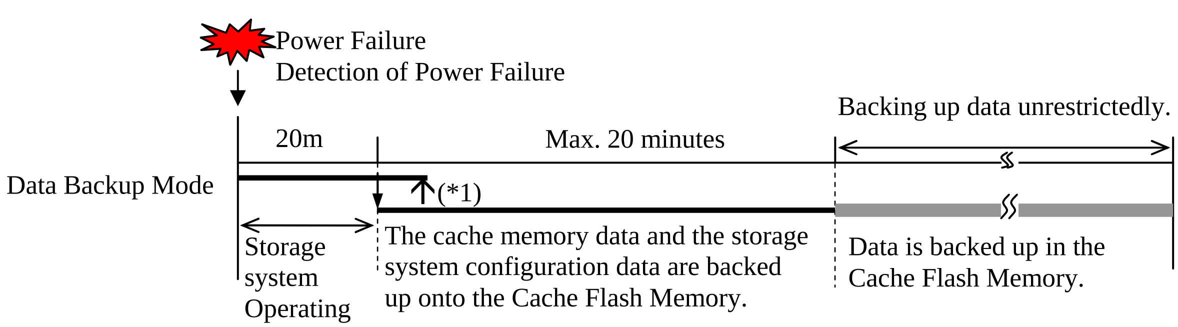
Model Number	Capacity	Component	Memory Chip Type
DF-F850-4GB	4GB	4GB DIMM ×1	DDR3 2Gbit 1Rank
DF-F850-8GB	8GB	8GB DIMM ×1	DDR3 2Gbit 2Rank
DW-F700-16GB	16GB	16GB DIMM ×1	DDR3 4Gbit 2Rank

(4) Cache Flash Memory (CFM)
The Cache Flash Memory saves the cache memory data when a power failure occurs.

Table 2.2.3-3 Cache Flash Memory Specifications

Model Number	Capacity	Component	Installation Requirements
DW-F700-BM160	160GB	80GB SSD ×2	The capacity of the maximum cache memory per MAIN Blades (two boards) is 160GB or less.
DW-F700-BM256	256GB	128GB SSD ×2	The capacity of the maximum cache memory per MAIN Blades (two boards) is 192GB or more. This option can be used when capacity of the cache memory per MAIN Blades is 160GB or less instead of DW-F700-BM160.

(5) Battery
The battery for the data saving is installed on each MAIN Blade board in DW700. When the power failure continues more than 20 milliseconds, the storage system uses power from the batteries to backup the cache memory data and the storage system configuration data onto the Cache Flash Memory. Environmentally friendly nickel hydride battery is used for the storage system.
DW700 supports neither the UPS function for a minute nor the de-stage function.



*1: The data backup processing is continued when the power outage recovered while the data is being backed up.

Fig. 2.2.3-5 Data Backup Process

 Copyright © 2013, Hitachi, Ltd.

(5-1) Relation between Battery Charge Level and System Startup Action

No.	Power Status	Battery Charge Level	System Startup Action
1	PS ON	<Case1> The battery charge level of both the MAIN Blades is below 25%.	The system does not start up until the battery charge level of both the MAIN Blades becomes 25% or more in either or both of the clusters. (It takes a maximum of 85 minutes (*2).) (*1)
2		<Case2> The battery charge level of both the MAIN Blades is below 50%. (In the case other than Case1)	SIM that shows the lack of battery charge is reported and the system starts up. I/O is executed by the pseudo through operation until the battery charge level of either or both of the MAIN Blades becomes 50% or more. (It takes a maximum of 50 minutes (*2).)
3		<Case3> Other than Case1, 2. (The battery charge level of either or both of the MAIN Blades is 50% or more.)	The system starts up normally. If the condition changed from Case2 to Case3 during startup, SIM that shows the completion of battery charge is reported.

*1: Action when System Option Mode 885 is off (default setting).
*2: Battery charge time: 4.5 hours to charge from 0% to 100%.

(5-2) Relation between Power Status and SM/CM Data Backup Methods

No.	Power Status		SM/CM Data Backup Methods	Data Restore Methods during Restart
1	PS OFF (planned power off)		SM data (including CM directory information) is stored in CFM before PS OFF is completed. If PIN data exists, all the CM data including PIN data is also stored.	SM data is restored from CFM. If CM data was stored, CM data is also restored from CFM.
2	When power outage occurs	Instant power outage	If power is recovered in a moment, SM/CM data remains in memory and is not stored in CFM.	SM/CM data in memory is used.
3		Power outage while the system is in operation	All the SM/CM data is stored in CFM. If a power outage occurred after the system started up in the condition of Case2 (the battery charge level of both the MAIN Blades had been below 50%), only SM data is stored.	All the SM/CM data is restored from CFM. If CM data was not stored, only CM data is volatilized and the system starts up.
4		Power outage while the system is starting up	Data storing in CFM is not done. (The latest backup data that was successfully stored remains.)	The data that was stored in the latest power off operation or power outage is restored from CFM.

Copyright © 2013, Hitachi, Ltd.

(5-3) Action When CFM Error Occurs

No.	DKC Status	Description of Error	Action When Error Occurs
1	In operation	CFM error or data comparing error was detected at the time of CFM health check (*1).	CFM warning SIMRC = BFD21X (Environmental error: CFM Warning) is generated.
2	Planed power off or power outage	CFM error was detected, and moreover, retry failed four times during data storing. Data storing error is managed in a per module group (MG) basis and is classified into data storing error only in the MG concerned and data storing error in all the MG depending on the location of the failed memory.	① DKC power off process is executed. ② Blockage occurs in MAIN Blade or CMG in MAIN Blade depending on the location of the failed memory. For details, see THEORY03-09-20.
3	When powered on -1 (In the case that data storing was successfully done in No.2)	CFM error or protection cord (*2) error occurred during data restoring.	① Blockage occurs in MAIN Blade or CMG in MAIN Blade depending on the location of the failed memory. ② If the failed memory is in CMG0, the MAIN Blade concerned becomes blocked. If the failed memory is in CMG1, the CACHE concerned is volatilized and the system starts up. (If data in the other MAIN Blade can be restored, the data is not lost.)
4	When powered on -2 (In the case that data storing failed in No.2)	—	① Blockage occurs in MAIN Blade or CMG in MAIN Blade depending on the location of data storing error. ② Same as No.3-②

*1: CFM health check: Function that executes the test of reading and writing of a certain amount of data at specified intervals to CFM while the DKC is in operation.
*2: Protection code: The protection code (CRC) is generated and saved onto CFM at the time of data storing in CFM and is checked at the time of data restoring.

NOTE: CFM handles only the data in the MAIN Blade in which it is installed.
e.g.: Cache data in MAIN1 is not stored in CFM which is installed in MAIN2.
Similarly, CFM data in the MAIN1 is not restored to cache memory in MAIN2.

(5-4) Notes during Planned Power Off (PS OFF)

Removing the MAIN Blade when the system is off and the breakers on the PDU are on may result in Case1 because of the lack of battery charge. Therefore, to remove the blade and the battery, replace them when the system is on, or remove them after the breakers on the PDU are powered off.

 Copyright © 2012, 2013, Hitachi, Ltd.

(6) MP Blades (MPB)

The MP Blade is a processor board that includes a chipset, DIMMs and Fans, controls Host I/O Module, Backend I/O Module, PCI-express and local memory, and communicates with SVP by Ethernet.

Table 2.2.3-4 MP Blades Specifications

PCB Name	WP790-A
Number of PCB	1
Necessary Number of PCB per Controller Chassis	2
Performance of Processor	2.1GHz
Processor Type	Xeon E5-2450 (8 cores)
Number of Processor per PCB	1
Local Memory Capacity per PCB	8GB (4GB DIMM × 2)

(7) Backend I/O Module (DKB)

The Backend I/O Module (DKB) controls data transfer between the drive and cache memory.

Table 2.2.3-5 Backend I/O Module Specifications

Model Number	DW-F700-BS6G	DW-F700-BS6GE
PCB Name	SH561-B	SH561-D
Number of PCB	1	1
Number of SAS Port per PCB	2	2
Performance of SAS Port	6Gbps/3Gbps	6Gbps/3Gbps
Data Encryption	Not Support	Support
Necessary Number of PCBs per Controller Chassis	4	4
Maximum Number of SAS Port per Storage System	8	8
Maximum number of Disk Drives per SAS Port	144 (2.5-inch HDD) 144 (3.5-inch HDD)	144 (2.5-inch HDD) 144 (3.5-inch HDD)

Copyright © 2012, Hitachi, Ltd.

(8) Host I/O Module (CHB)

The Host I/O Module (CHB) controls data transfer between the upper host and the cache memory.

It is necessary to install or add the Host I/O Module (CHB) two pieces at a time.

Table 2.2.3-6 CHB Specifications

		Fibre 8Gbps
Model Number		DF-F850-HF8GR (*2)
PCB Name		SH559-C
Number of PCB		1
Host Interface		FCP
Data Transfer Rate (MB/s)		200/400/800
Number of Options Installed ( ): DKB slot used		2/4/6/8 (10/12)
Number of Ports per Option		4
Number of Ports per storage system ( ): DKB slot used		8/16/24/32 (40/48)
Maximum Cable Length	Shortwave	500m/380m/150m (*1)
	Longwave	10km

*1: When 50/125µm laser optimized multi-mode fiber cable (OM3) is used. When using other cable types, it is limited to the length shown in Table 2.2.3-7.

*2: The CHB for the fibre channel connection can conform to either short wavelength or long wavelength by selecting a transceiver installed on each port. However, DW-F700-1UL (SFP for 8Gbps Longwave) must be installed for the long wavelength, because the short wavelength transceiver is included in the CHB as a standard.

Table 2.2.3-7 Maximum cable length (Shortwave)

Data Transfer Rate	Maximum cable length		
	OM1 (62.5/125µm multi-mode fiber)	OM2 (50/125µm multi-mode fiber)	OM3 (50/125µm laser optimized multi-mode fiber)
200MB/s	150m	300m	500m
400MB/s	70m	150m	380m
800MB/s	21m	50m	150m

Copyright © 2012, Hitachi, Ltd.

(9) Service Processor (SVP)

The Service Processor (SVP) mainly performs a setting and a modification of the storage system configuration, a device availability statistical information acquisition, and maintenance.
The modem card is not used in DW700.
A dual SVP configuration is not supported in DW700.
Maintenance switches are prepared on a front interface of SVP (for CE maintenance personnel use only).

Table 2.2.3-8 SVP Specifications

Item		Specifications
OS		Windows Vista Business
Keyboard		— (*1)
Display		— (*1)
LAN		On-Board 10Base-T / 100Base-TX / 1000Base-T × 3 Port
Device	FDD	None
	CD-ROM	None
Serial Port		None
USB		Version 1.1 / 2.0 × 4
PC Card Slot		None

*1: The maintenance PC (Console PC) with exact specification must be prepared and connects with SVP, to implement the installation or maintenance of the storage system, because the exclusive SVP for DW700 has neither a display nor a keyboard.

Table 2.2.3-9 Maintenance PC (Console PC) Specifications (*1)

Item	Specifications
OS	Windows XP / Windows Vista / Windows 7
Disk Drive	Available hard disk space: 500MB or more
Display	1024 × 768 (XGA) or higher-resolution 1280 × 1024 (SXGA) Recommendation
CD-ROM	Need
LAN	Ethernet 1000Base-T, 10Base-T/100Base-TX
USB	Need
Requisite software	Adobe Reader 8 or higher

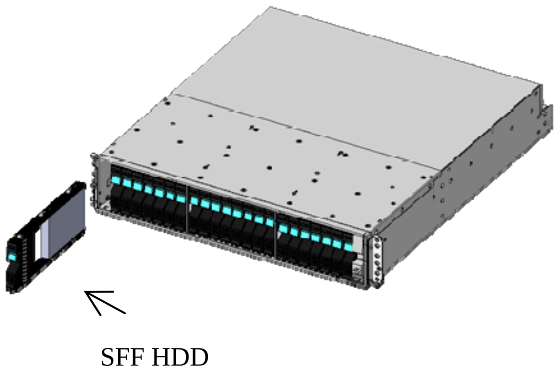
*1: A power outlet is required to use the maintenance PC (Console PC). When connecting with SVP, the keyboard layout on the maintenance PC (Console PC) will be English 101/104-based layout.

Copyright © 2012, 2013, Hitachi, Ltd.

(10) Drive Box (DBS)

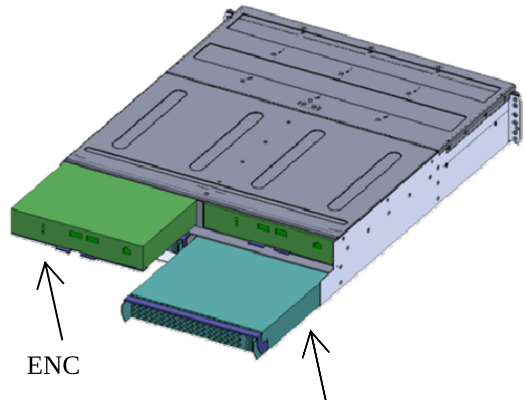
The Drive Box (DBS) is a chassis to install the 2.5-inch disk drives and the 2.5-inch flash drives, and consists of two ENCs and two power supplies with a built-in cooling fan.

Front View



24 SFF HDDs can be installed.

Rear View



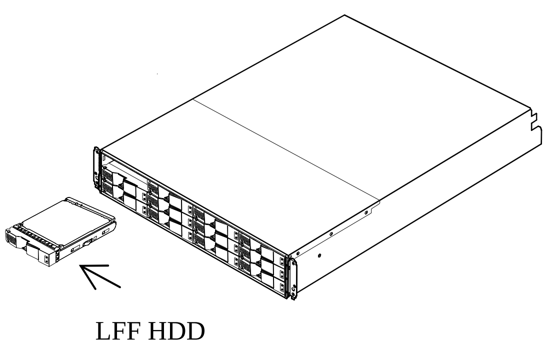
ENC and power supply are a duplex configuration.

Fig. 2.2.3-6 Drive Box (DBS)

(11) Drive Box (DBL)

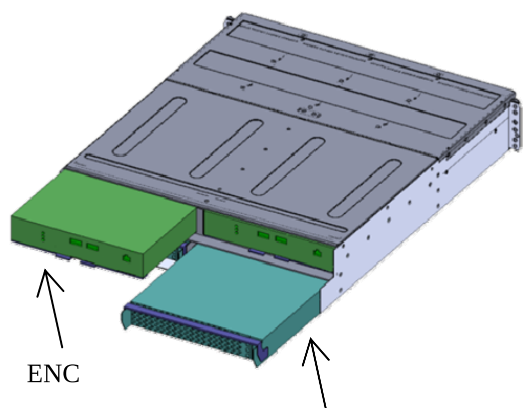
The Drive Box (DBL) is a chassis to install the 3.5-inch disk drives and the 3.5-inch flash drives, and consists of two ENCs and two power supplies with a built-in cooling fan.

Front View



12 LFF HDDs can be installed.

Rear View



ENC and power supply are a duplex configuration.

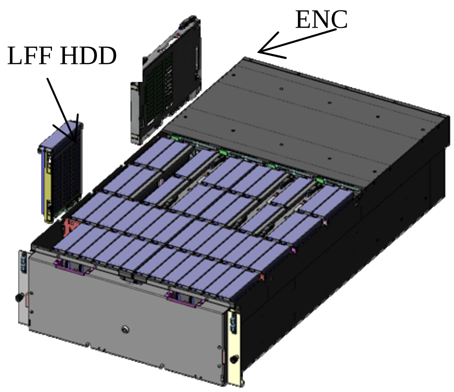
Fig. 2.2.3-7 Drive Box (DBL)

 Copyright © 2012, 2013, Hitachi, Ltd.

(12) Drive Box (DBX)

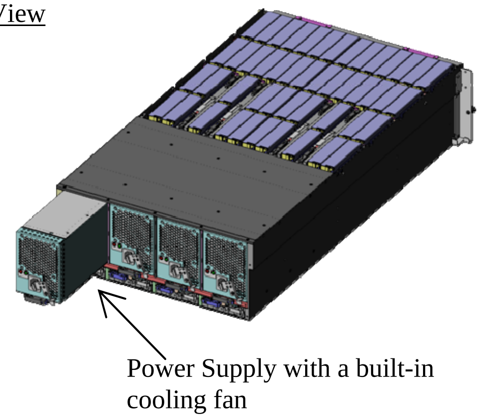
The Drive Box (DBX) is a chassis to install the 3.5-inch disk drives, and consists of four ENCs, four power supplies with a built-in cooling fan.

Front View



48 LFF HDDs can be installed.

Rear View



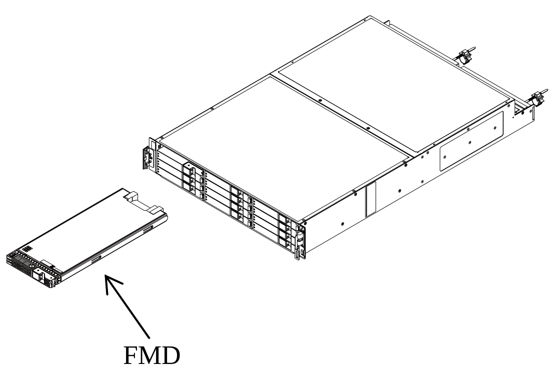
Drive Box (DBX) is configured with 2 units, 24 of LFF drives can be installed in each unit.
The ENCs and power supplies are individual for each unit, and are adopted duplex configuration.

Fig. 2.2.3-8 Drive Box (DBX)

(13) Drive Box (DBF)

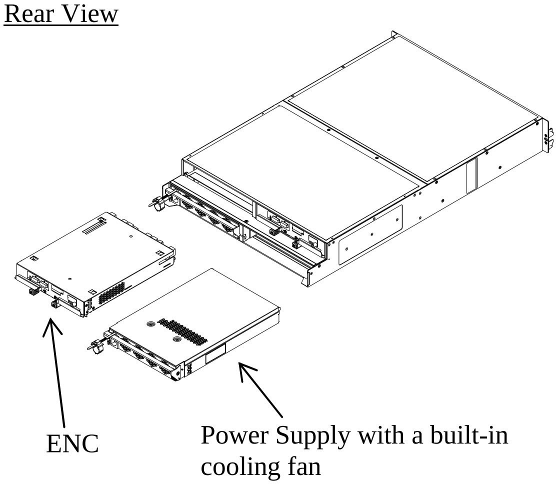
The Drive Box (DBF) is a chassis to install the flash module drives, and consists of two ENCs and two power supplies with a built-in cooling fan.

Front View



12 FMDs can be installed.

Rear View



ENC and power supply are a duplex configuration.

Fig. 2.2.3-9 Drive Box (DBF)

Copyright © 2012, 2013, Hitachi, Ltd.

(14) Disk Drive, Flash Drive and Flash Module Drive

The disk drive, the flash drive and the flash module drive supported by DW700 are shown below.

Table 2.2.3-10 Disk Drive and Flash Drive Support Type

Group	I/F	Size (inch)	Transfer Rate (Gbps)	Revolution Speed (min^{-1}) or Flash Memory Type	Capacity
Disk Drive	SAS	2.5	6	15,000	300GB
		2.5	6	10,000	600GB, 900GB, 1.2TB
		3.5	6	7,200	3TB, 4TB
Flash Drive	SAS	2.5	6	MLC	200GB, 400GB, 800GB
Flash Module Drive	SAS	—	6	MLC	1.6TB, 3.2TB

Table 2.2.3-11 LFF Disk Drive Specifications

Item		3TB			4TB
Disk Drive Model Name	Seagate	S2D-H3R0SS	—	S2E-H3R0SS	S2E-H4R0SS
	HGST	—	R2D-H3R0SS	—	—
Capacity (GB)		2,937.11	2,937.11	2,937.11	3916.14
Number of heads		10	10	8	10
Number of disks		5	5	4	5
Seek Time (ms) (Read/Write)	Average (*1)	8.1/8.8	8.2/9.2	7.8/8.5	7.8/8.5
Average latency time (ms)		4.16	4.2	4.16	4.16
Revolution speed (min^{-1})		7,200	7,200	7,200	7,200
Interface data transfer rate (Gbps)		6	6	6	6
Internal data transfer rate (MB/s)		Max. 239	97.9 to 203.9	Max. 276	Max. 276

*1: Not including controller overhead

Copyright © 2012, 2013, Hitachi, Ltd.

Table 2.2.3-12 SFF Disk Drive Specifications

Item		300GB	600GB		
Disk Drive Model Name	Seagate	S5C-K300SS	S5D-J600SS	—	—
	HGST	—	—	R5C-J600SS	R5D-J600SS
Capacity (GB)		288.20	576.39	576.39	576.39
Number of heads		4	4	6	4
Number of disks		2	2	3	2
Seek Time (ms) (Read/Write)	Average (*1)	2.7/3.1	3.4/3.8	3.7/4.2	3.8/4.2
Average latency time (ms)		2.0	3.0	3.0	3.0
Revolution speed (min^{-1})		15,000	10,000	10,000	10,000
Interface data transfer rate (Gbps)		6	6	6	6
Internal data transfer rate (MB/s)		194.3 to 283	125 to 238.8	152.4 to 253.6	164.9 to 279

Item		600GB	900GB		
Disk Drive Model Name	Seagate	S5E-J600SS	S5D-J900SS	—	S5E-J900SS
	HGST	—	—	R5D-J900SS	—
Capacity (GB)		576.39	864.64	864.64	864.64
Number of heads		4	6	6	6
Number of disks		2	3	3	3
Seek Time (ms) (Read/Write)	Average (*1)	3.6/4.1	3.7/4.1	3.9/4.2	3.6/4.1
Average latency time (ms)		2.9	3.0	3.0	2.9
Revolution speed (min^{-1})		10,000	10,000	10,000	10,000
Interface data transfer rate (Gbps)		6	6	6	6
Internal data transfer rate (MB/s)		180 to 293.8	125 to 238.8	164.9 to 279	180 to 293.8

Item		1.2TB			
Disk Drive Model Name	Seagate	—			
	HGST	R5E-J1R2SS			
Capacity (GB)		1152.79GB			
Number of heads		8			
Number of disks		4			
Seek Time (ms) (Read/Write)	Average (*1)	4.6/5.0			
Average latency time (ms)		3.0			
Revolution speed (min^{-1})		10,000			
Interface data transfer rate (Gbps)		6			
Internal data transfer rate (MB/s)		161.1 to 279.0			

*1: Not including controller overhead

 Copyright © 2012, 2013, Hitachi, Ltd.

Table 2.2.3-13 SFF Flash Drive Specifications

Item		200GB		400GB	
Flash Drive Model Name	HGST	R5B-M200SS	—	R5B-M400SS	—
	Toshiba	—	B5A-M200SS	—	B5A-M400SS
Form Factor		2.5-inch	2.5-inch	2.5-inch	2.5-inch
Capacity (GB)		196.92	196.92	393.85	393.85
Flash memory technology		MLC	MLC	MLC	MLC
Interface data transfer rate (Gbps)		6	6	6	6

Item		800GB			
Flash Drive Model Name	HGST	—			
	Toshiba	B5A-M800SS			
Form Factor		2.5-inch			
Capacity (GB)		787.69			
Flash memory technology		MLC			
Interface data transfer rate (Gbps)		6			

Table 2.2.3-14 Flash Module Drive Specifications

Item	1.6TB	3.2TB		
Flash Module Drive Model Name	H1A-P1R6SS	H1B-P3R2SS		
Form Factor	—	—		
Capacity (GB)	1,759.2	3,518.4		
Flash memory technology	MLC	MLC		
Interface data transfer rate (Gbps)	6	6		

Copyright © 2013, Hitachi, Ltd.

(15) FMD

FMD has 27.75Wh Li-ion battery in it.

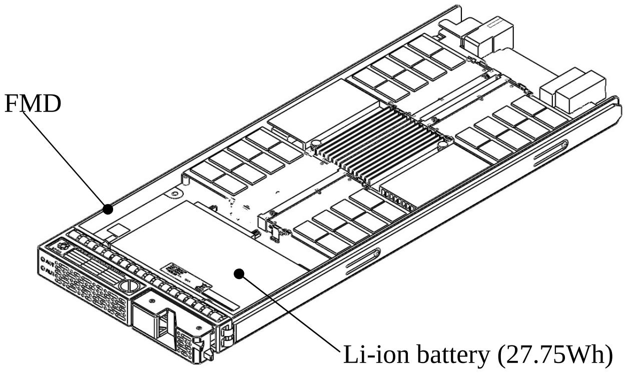


Fig. 2.2.3-10 Location of Li-ion Battery

Package with over 100Wh battery should be handled as DG (Dangerous Goods) in IATA regulation when it is transported by airplane. DG freight requires additional freight fee. So, if you ship multiple FMDs installed in DBF, it will be handled as DG.
In order not to be handled as DG, please transport FMD in single module package.
If you ship FMDs installed in DBF as DG, you have to display “DG Mark label” below in outer package. And the package must be the one specified by U. N, but the present package does not satisfy it.

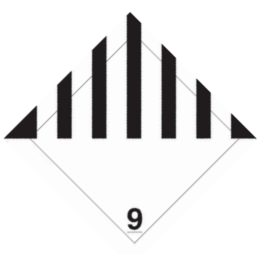


Fig. 2.2.3-11 DG Mark label

Copyright © 2013, Hitachi, Ltd.

Even if it is the case that multiple single module packages are put into one exterior package and it is transported by air, you have to display “Battery label” below in outer package.

Fig. 2.2.3-12 Battery label

However, the interpretation of IATA regulation is a little different by each airline companies. So, although it depends on the number of FMDs, even if FMDs are installed in DBF, when using the airline which does not treat as a DG article, then the treatment in this clause is unnecessary.
Also, even if multiple single module packages are put into one exterior package, when using the airline which does not treat the package as “Battery label” necessary article, then the treatment in this clause is unnecessary.

Copyright © 2012, 2013, Hitachi, Ltd.

2.3 Storage system Specifications

DW700 storage system specifications are shown in the following table.

Table 2.3-1 Storage system specifications

<table>
<tr><th colspan="4">Item</th><th>Specifications</th></tr>
<tr><td rowspan="12">System</td><td rowspan="2" colspan="2">Number of Disk Drives</td><td>Minimum</td><td>4 (disk-in model) / 0 (diskless model)</td></tr>
<tr><td>Maximum</td><td>1,152</td></tr>
<tr><td colspan="3">Maximum Number of Flash Drives</td><td>128 (*1)</td></tr>
<tr><td colspan="3">Maximum Number of Flash Module Drives</td><td>96</td></tr>
<tr><td colspan="3">RAID Level</td><td>RAID6/RAID5/RAID1</td></tr>
<tr><td rowspan="3" colspan="2">RAID Group Configuration</td><td>RAID6</td><td>6D+2P, 14D+2P</td></tr>
<tr><td>RAID5</td><td>3D+1P, 7D+1P</td></tr>
<tr><td>RAID1</td><td>2D+2D, 4D+4D</td></tr>
<tr><td colspan="3">Maximum Number of Spare Drives</td><td>64 (*2)</td></tr>
<tr><td colspan="3">Maximum Number of LDEVs</td><td>16,384</td></tr>
<tr><td colspan="3">Maximum Storage System Capacity (Physical Capacity)</td><td>4,511TB (4TB SAS HDD used)</td></tr>
<tr><td colspan="3" style="display:none"></td></tr>
<tr><td rowspan="2">Memory</td><td colspan="3">Cache Memory Capacity</td><td>32GB/64GB/96GB/128GB/160GB/192GB/256GB</td></tr>
<tr><td colspan="3">Cache Flash Memory Capacity</td><td>160GB/256GB</td></tr>
<tr><td rowspan="4">Device I/F</td><td colspan="3">DKC-DB Interface</td><td>SAS/Dual Port</td></tr>
<tr><td colspan="3">Data Transfer Rate</td><td>6Gbps</td></tr>
<tr><td colspan="3">Maximum Number of HDD per SAS I/F</td><td>144</td></tr>
<tr><td colspan="3">Number of DKB PCB</td><td>4</td></tr>
<tr><td rowspan="3">Channel I/F</td><td colspan="3">Support Channel Type</td><td>Fibre Channel Short Wavelength (*3)</td></tr>
<tr><td colspan="2">Data Transfer Rate (MB/s)</td><td>Fibre Channel</td><td>200 / 400 / 800</td></tr>
<tr><td colspan="3">Maximum Number of CHB</td><td>8 (12: DKB Slot used)</td></tr>
<tr><td>Power</td><td colspan="2">AC Input</td><td>Single Phase</td><td>60Hz : 200V to 240V
50Hz : 200V to 240V</td></tr>
<tr><td rowspan="4">Acoustic Level (*4)</td><td rowspan="2">Operating</td><td colspan="2">-CBX</td><td>60dB</td></tr>
<tr><td colspan="2">-DBL/DBS/DBX/DBF</td><td>60dB/60dB/62dB/60dB</td></tr>
<tr><td rowspan="2">Standby</td><td colspan="2">-CBX</td><td>55dB</td></tr>
<tr><td colspan="2">-DBL/DBS/DBX/DBF</td><td>55dB/55dB/57dB/55dB</td></tr>
</table>

(To be continued)

Copyright © 2012, 2013, Hitachi, Ltd.

(Continued from preceding page)

Item		Specifications
Non Stop Maintenance	Control PCB	Support
	Cache Memory Module	Support
	Cache Flash Memory	Support
	Power Supply, Fan	Support
	Microcode	Support
	Disk Drive, Flash Drive, Flash Module Drive	Support

*1: The number of spare drives is not included. Spare drives are installed in the same slots as those for spare HDDs.
*2: Available as spare or data disks.
*3: By the replacing SFP transceiver of the fibre port on the CHB to the DW-F700-1UL, the port can be used for the long wavelength.
*4: Measurement Condition: The point 1m far from floor and surface of the product.

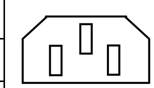
 Copyright © 2012, 2013, Hitachi, Ltd.

2.4 Power Specifications

2.4.1 Storage system Power Specifications

DW700 input power specifications are shown as each power supply.

Table 2.4.1-1 Input Power Specifications

Item	Input Power	Input Current (*1)	Steady Current (*2)	Leakage Current	Inrush Current			Power Cord Plug Type
					1st (0-p)	2nd (0-p)	1st (0-p) Time	
DKC PS	1-phase, AC200V to AC240V	3.81A	1.91A	0.28mA	25A	20A	150ms	IEC60320 C14 
DBL PS		2.07A	1.04A	1.75mA	25A	20A	150ms	
DBS PS		2.61A	1.31A	1.75mA	25A	20A	150ms	
DBX PS		4.03A	2.02A	0.87mA	30A	30A	150ms	
DBF PS		2.6A	1.3A	0.28mA	20A	15A	80ms	

*1: The maximum current in case AC input is not a redundant configuration (in case of 184V [200V -8%]).
*2: The maximum current in case AC input is a redundant configuration (in case of 184V [200V -8%]).

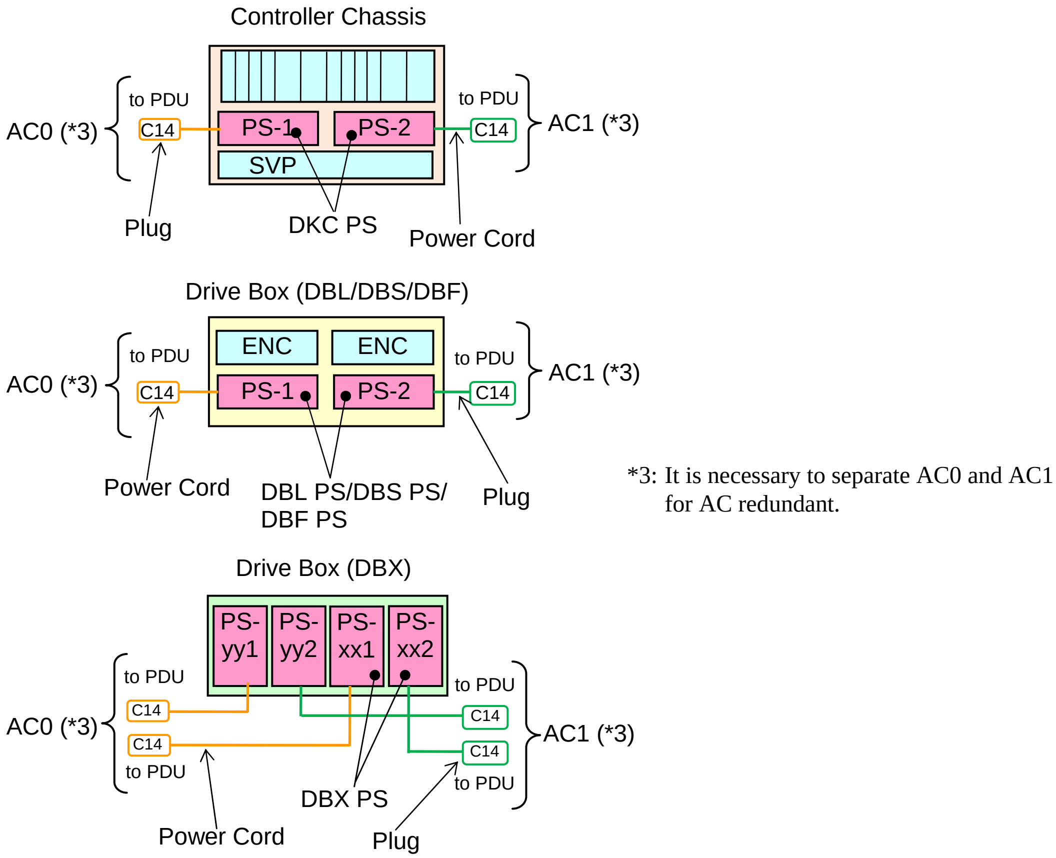


Fig. 2.4.1-1 Power Supply Locations

Copyright © 2012, 2013, Hitachi, Ltd.

2.5 Environmental Specifications

The environmental specifications are shown in the following table.

Table 2.5-1 Environmental Specifications

Item	Condition		
	Operation (*1)	Non-operation (*2)	Transportation, Storage (*3)
Temperature range (ºC)	10 to 40 (*5) 10 to 35 (*6)	-10 to 50 -10 to 35 (*7)	-30 to 60 -10 to 35 (*7)
Relative humidity (%) (*4)	8 to 80	8 to 80	5 to 95
Maximum wet-bulb temperature (ºC)	29	29	29
Temperature gradient (ºC/hour)	10	10	10
Dust	Below 0.15mg/m^3	—	—
Altitude	-60m to 3,000m	-60m to 12,000m	-60m to 12,000m

*1: Environmental conditions of operation should be completed before switch on a system.
*2: “Non-operation” includes conditions of both packing and unpacking.
*3: Transportation and storage should be conducted in the packing of initial shipping.
*4: No dew condensation.
*5: For Controller Chassis and Drive Box (SFF/LFF).
*6: For Drive Box (DENSE) and Flash Module Drive Box.
*7: For Flash Module Drive (for DBF).

Table 2.5-2 Mechanical Environmental Specifications

Item	In operation	In non-operation
Guaranteed value to vibration	Below 2.45m/s^2 (0.25G)	Below 3.9m/s^2 (0.4G) : No critical damage for product function. (Normal operating with part replacement)
		Below 9.8m/s^2 (1.0G) : Ensure own safety with fall prevention.
Guaranteed value to impact	No impact	78.4m/s^2 (8.0G), 15ms
Guaranteed value to seismic wave	Below 2.45m/s^2 (0.25G) (250gal approx.)	Below 3.9m/s^2 (0.4G) (400gal) : No critical damage for product function. (Normal operating with part replacement) Below 9.8m/s^2 (1.0G) (1000gal) : Ensure own safety with fall prevention.

Copyright © 2012, 2013, Hitachi, Ltd.

3. Internal Operation

3.1 Hardware Block Diagram

The system of DW700-RK consists of Controller Chassis (DKC) and Drive Box (DB). There are four kinds of DB (2.5 inches DBS, 3.5 inches DBL, DBF and 3.5 inches DBX).

When DBS or DBL is mounted, one DKC and up to 16 DBs can be installed in one rack. In case of maximum configuration, 48 DBs can be installed on the system. When DBX is mounted, one DKC and up to 5 DBXs can be installed in one rack. In case of maximum configuration, 24 DBs can be installed on the system.
Figure 3.1-1 shows the composition in the rack. DBS, DBL, DBF and DBX can exist together in the same system.

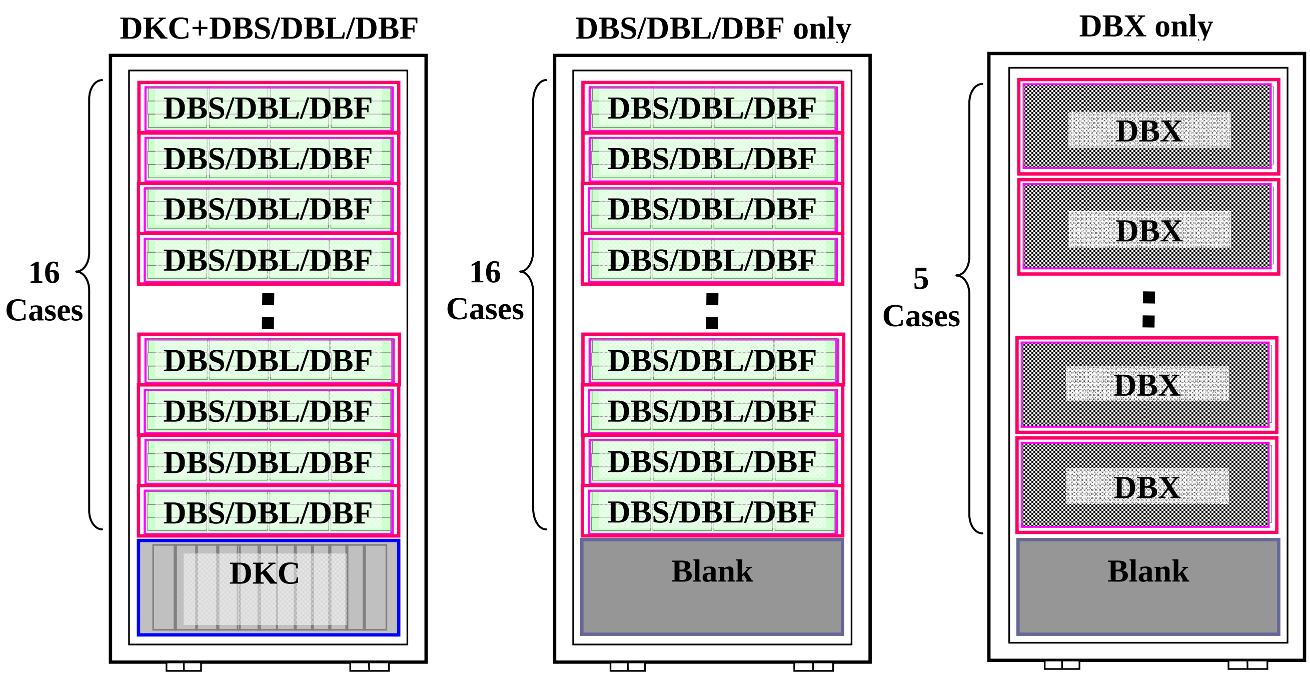


Fig. 3.1-1 Composition in the rack

Copyright © 2012, Hitachi, Ltd.

The controller chassis (DKC) consists of MAIN Blades, processor blades (MPB), Channel blades (CHB), disk blades (DKB), service processor (SVP), Cooling Fan and AC-DC power supply that supplies the power to the components.
The Drive Chassis is a chassis to install the disk drives and the flash drives, and consists of ENC board, and the AC-DC power supply with the cooling fan.

Figure 3.1-1 shows the hardware block diagram of storage system.

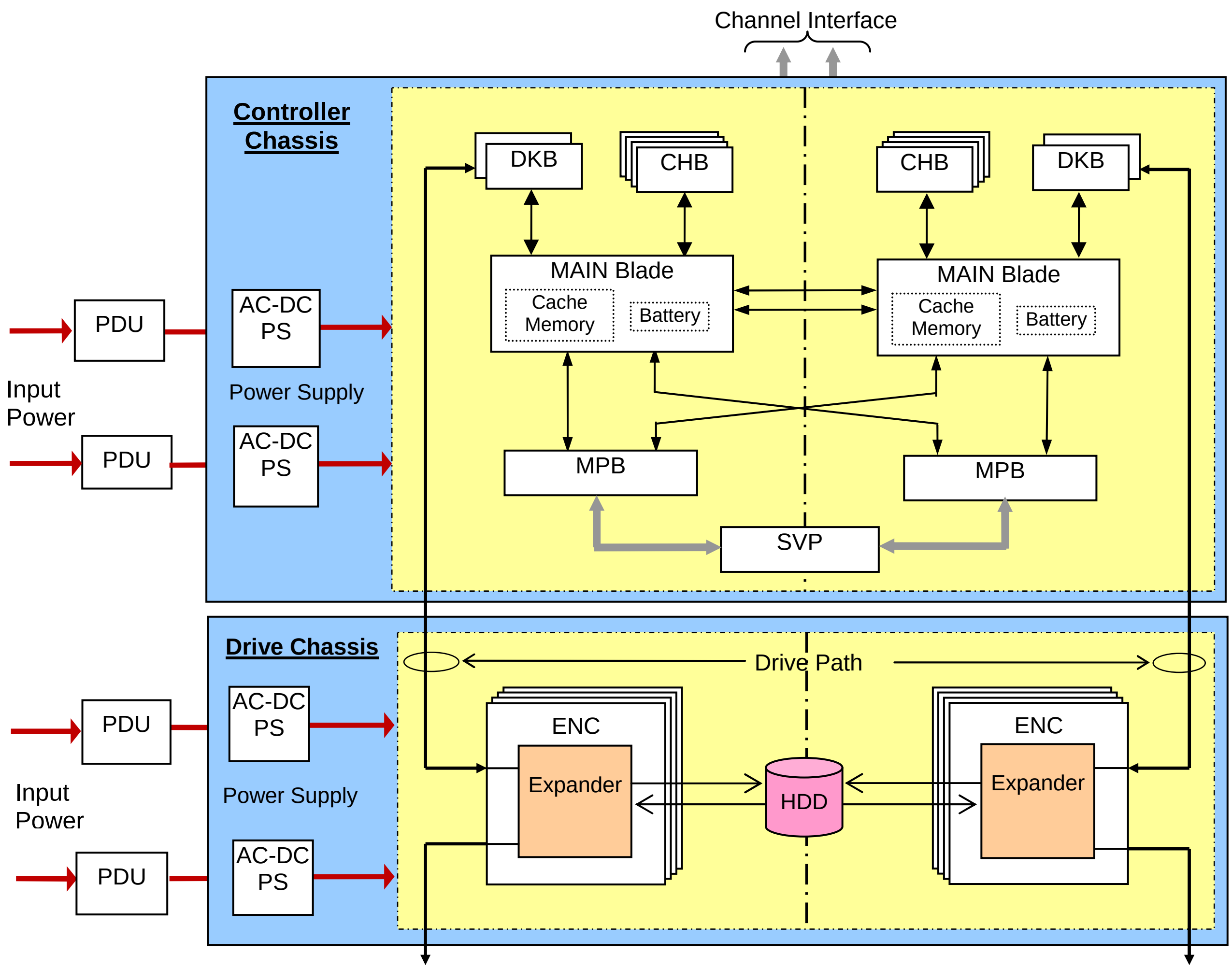


Fig. 3.1-2 Hardware block diagram of Storage system

3.2 Software Organization

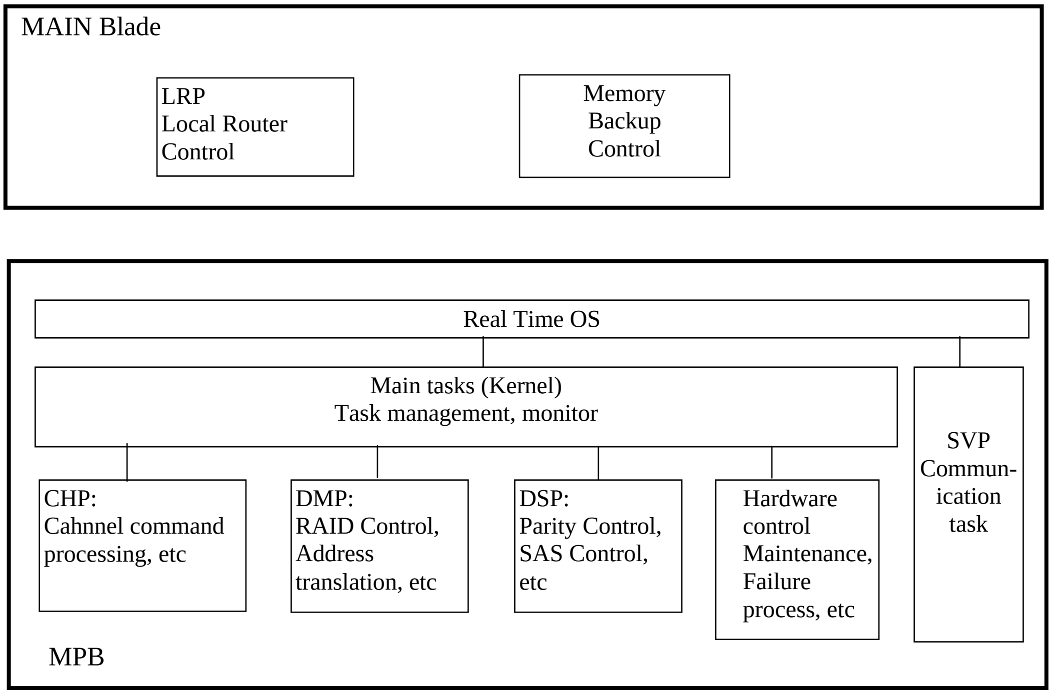


Fig. 3.2-1 Software Organization

Real Time OS:

A basic OS for controlling the RISC processor. Its primary tasks are to control and switch between the main tasks and SVP communication tasks.

Main tasks:

Made up of DKC control tasks (CHP, DMP, DSP) and the kernel tasks that supervise the DKC control tasks. They switch the control tasks by making use of the kernel's task switching facility.

SVP communication task:

Controls the communication with the SVP.

CHP (CHannel Program):

Is a channel command control layer that processes channel commands and controls cache and data transfer operations. CHP is recognized by the logical volume number and logical block number.

Copyright © 2012, Hitachi, Ltd.

DMP (Disk Master Program):

RAID control functions.DMP is recognized by the logical volume number and logical block number.

DSP (Disk Slave Program):

Is a drive control layer and provides SAS control, drive data transfer control, and parity control functions. DSP is recognized by the physical volume number and LBA number.

 Copyright © 2012, Hitachi, Ltd.

3.3 Cache Management

Since the DKC requires no through operation, its cache system is implemented by two memory areas called cache A and cache B so that write data can be duplexed. To prevent data loss due to power failures, cache is made non-volatile by storing SSD on cache PCB. This dispenses with the need for the conventional NVS.

The minimum unit of cache is the segment. Cache is destaged in segment units. Emulation Disk type at one or four segments make up one slot. The read and write slots are always controlled in pair. Cache data is enqueued and dequeued usually in slot units. In real practice, the segments of the same slot are not always stored in a contiguous area in cache, but are stored in discreet areas. These segments are controlled suin-g CACHE-SLCB and CACHE-SGCB so that the segments belonging to the same slot are seemingly stored in a contiguous area in cache.

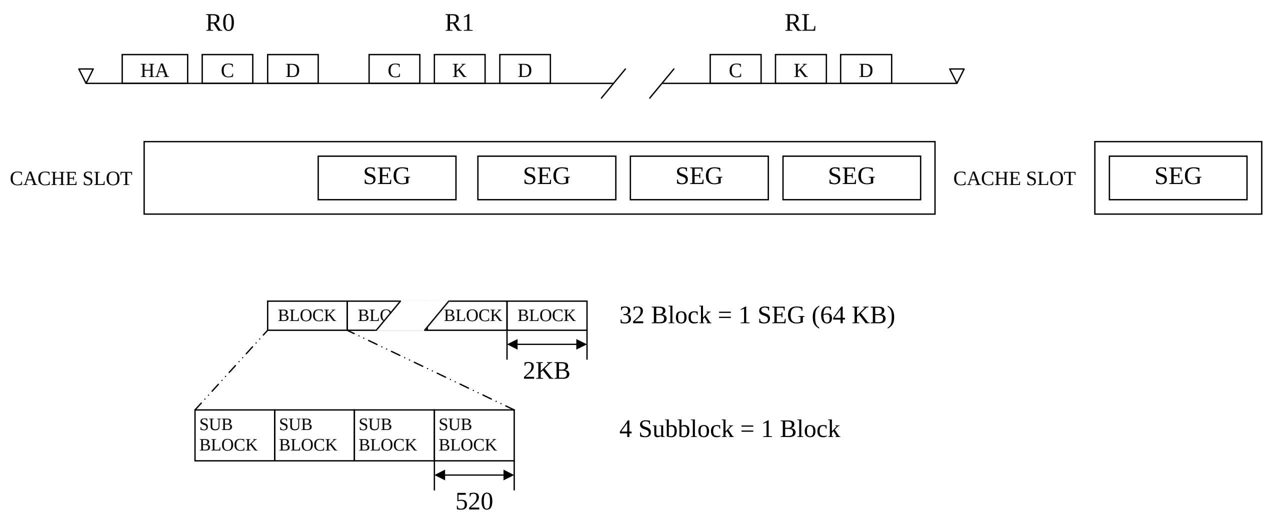


Fig. 3.3-1 Cache Data Structure

For increased directory search efficiency, a single virtual device (VDEV) is divided into 16-slot groups which are controlled using VDEV-GRPP and CACHE-GRPT.

1 cache segment = 32 blocks = 128 subblocks = 64 KB

1 slot = 1 stripe = 4 segments = 256 KB

The directories VDEV-GRPP, CACHE-GRPT, CACHE-SLCB, and CACHE-SGCB are used to identify the cache hit and miss conditions. These control tables are stored in the shared memory.

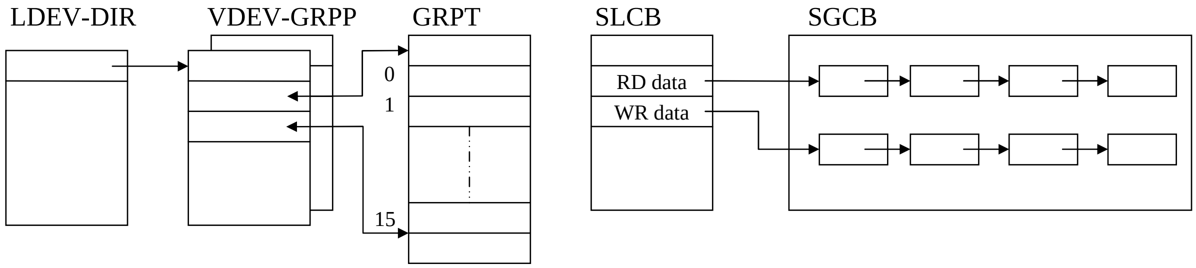
Copyright © 2012, Hitachi, Ltd.

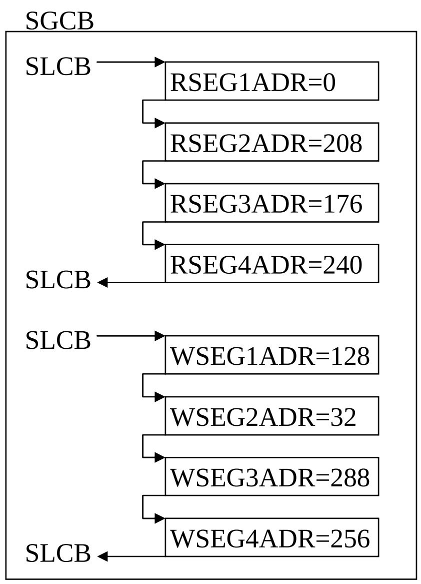
In addition to the cache hit and miss control, the shared memory is used to classify and control the data in cache according to its attributes. Queues are something like boxes that are used to classify data according to its attributes.

Basically, queues are controlled in slot units (some queues are controlled in segment units). Like SLCB-SGCB, queues are controlled using a queue control table so that queue data of the seemingly same attribute can be controlled as a single data group. These control tables are briefly described below.

Copyright © 2012, Hitachi, Ltd.

(1) Cache control tables (directories)





CACHE

	0	16	32	48
64	RSEG1		WSEG2	
128	WSEG1			RSEG3
192		RSEG2		RSEG4
256	WSEG4		WSEG3	
320				

LDEV-DIR (Logical DEV-directory):
Contains the shared memory addresses of VDEV-GRPPs for an LDEV. LDEV-DIR is located in the local memory in the CHB.

VDEV-GRPP (Virtual DEV-group Pointer):
Contains the shared memory addresses of the GRPTs associated with the group numbers in the VDEV.

GRPT (Group Table):
A table that contains the shared memory address of the SLCBs for 16 slots in the group. Slots are grouped to facilitate slot search and to reduce the space for the directory area.

SLCB (Slot Control Block):
Contains the shared memory addresses of the starting and ending SGCBs in the slot. One or more SGCBs are chained. The SLCB also stores slot status and points to the queue that is connected to the slot. The state transitions of clean and dirty queues occur in slot units. The processing tasks reserve and release cache areas in this unit.

SGCB (Segment Control Block):
Contains the control information about a cache segment. It contains the cache address of the segment. It is used to control the staged subblock bit map, dirty subblock bitmap, and other information. The state transitions of only free queues occur in segment units.

Copyright © 2012, Hitachi, Ltd.

(2) Cache control table access method (hit/miss identification procedure)

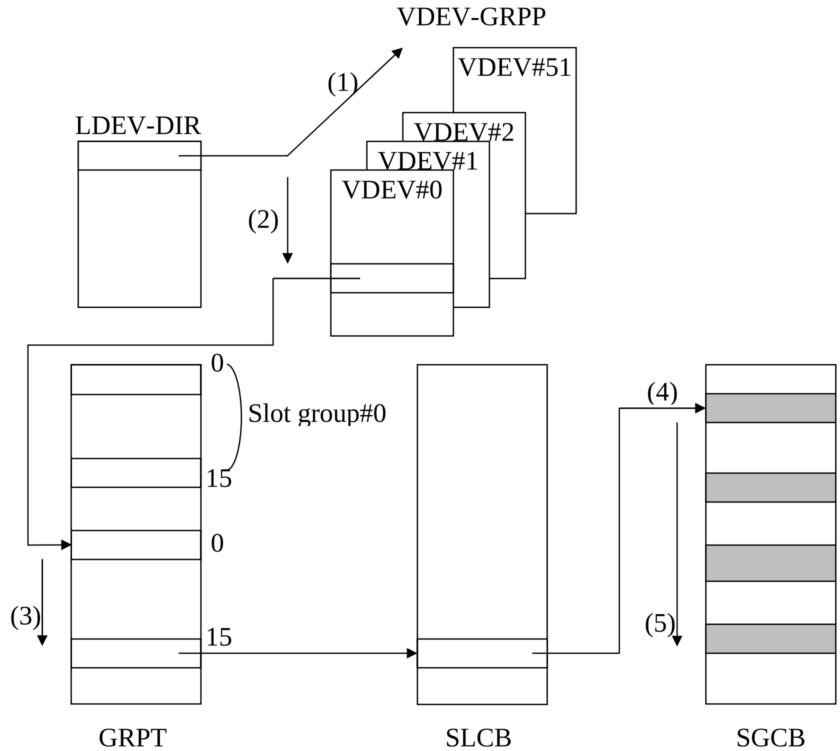


Fig. 3.3-2 Outline of Cache Control Table Access

1. The current VDEV-GRPP is referenced through the LDEV-DIR to determine the hit/miss condition of the VDEV-groups.
2. If a VDEV-group hits, CACHE-GRPT is referenced to determine the hit/miss condition of the slots.
3. If a slot hits, CACHE-SLCB is referenced to determine the hit/miss condition of the segments.
4. If a segment hits, CACHE-SGCB is referenced to access the data in cache.

If a search miss occurs during the searches from 1. through 4., the target data causes a cache miss.

Definition of VDEV number
Since the host processor recognizes addresses only by LDEV, it is unaware of the device address of the parity device. Accordingly, the RAID system is provided with a VDEV address which identifies the parity device associated with an LDEV. Since VDEVs are used to control data devices and parity devices systematically, their address can be computed using the following formulas:
Data VDEV number = LDEV number
Parity VDEV number = 1024 + LDEV number

From the above formulas, the VDEV number ranges from 0 to 2047.

 Copyright © 2012, Hitachi, Ltd.

(3) Queue structures

The DKC and DB uses 10 types of queues to control data in cache segments according to its attributes. These queues are explained below.

- CACHE-GRPT free queue

  This queue is used to control segments that are currently not used by CACHE-GRPT (free segments) on an FIFO (First-In, First-Out) basis. When a new table is added to CACHE-GRPT, the segment that is located by the head pointer of the queue is used.

- CACHE-SLCB free queue

  This queue is used to control segments that are currently not used by CACHE-SLCB (free segments) on an FIFO basis. When a new slot is added to CACHE-SLCB, the segment that is located by the head pointer of the queue is used.

- CACHE-SGCB free queue

  This queue is used to control segments that are currently not used by CACHE-SGCB (free segments) on an FIFO basis. When a new segment is added to CACHE-SGCB, the segment that is located by the head pointer of the queue is used.

- Clean queue

  This queue is used to control the segments that are reflected on the drive on an LRU basis.

- Bind queue

  This queue is defined when the bind mode is specified and used to control the segments of the bind attribute on an LRU basis.

- Error queue

  This queue controls the segments that are no longer reflected on the drive due to some error (pinned data) on an LRU basis.

- Parity in-creation queue

  This queue controls the slots (segments) that are creating parity on an LRU basis.

- DFW queue (host dirty queue)

  This queue controls the segments that are not reflected on the drive in the DFW mode on an LRU basis.

- CFW queue (host dirty queue)

  This queue controls the segments that are not reflected on the drive in the CFW mode on an LRU basis.

- PDEV queue (physical dirty queue)

  This queue controls the data (segments) that are not reflected on the drive and that occur after a parity is generated. Data is destaged from this queue onto the physical DEV. There are 32 PDEV queues per physical DEV.

  The control table for these queues is located in the shared memory and points to the head and tail segments of the queues.

Copyright © 2012, Hitachi, Ltd.

(4) Queue state transitions

Fig. 3.3-3 shows the state transitions of the queues used in. A brief description of the queue state transitions follows.

- State transition from a free queue

  When a read miss occurs, the pertinent segment is staged and enqueued to a clean queue. When a write miss occurs, the pertinent segment is temporarily staged and enqueued to a host dirty queue.

- State transition from a clean queue

  When a write hit occurs, the segment is enqueued to a host dirty queue. Transition from clean to free queues is performed on an LRU basis.

- State transition from a host dirty queue

  The host dirty queue contains data that reflects no parity. When parity generation is started, a state transition occurs to the parity in-creation queue.

- State transition from the parity in-creation queue

  The parity in-creation queue contains parity in-creation data. When parity generation is completed, a transition to a physical dirty queue occurs.

- State transition from a physical dirty queue

  When a write hit occurs in the data segment that is enqueued in a physical dirty queue, the segment is enqueued into the host dirty queue again. When destaging of the data segment is completed, the segment is enqueued into a queue (destaging of data segments occur asynchronously on an LRU basis).

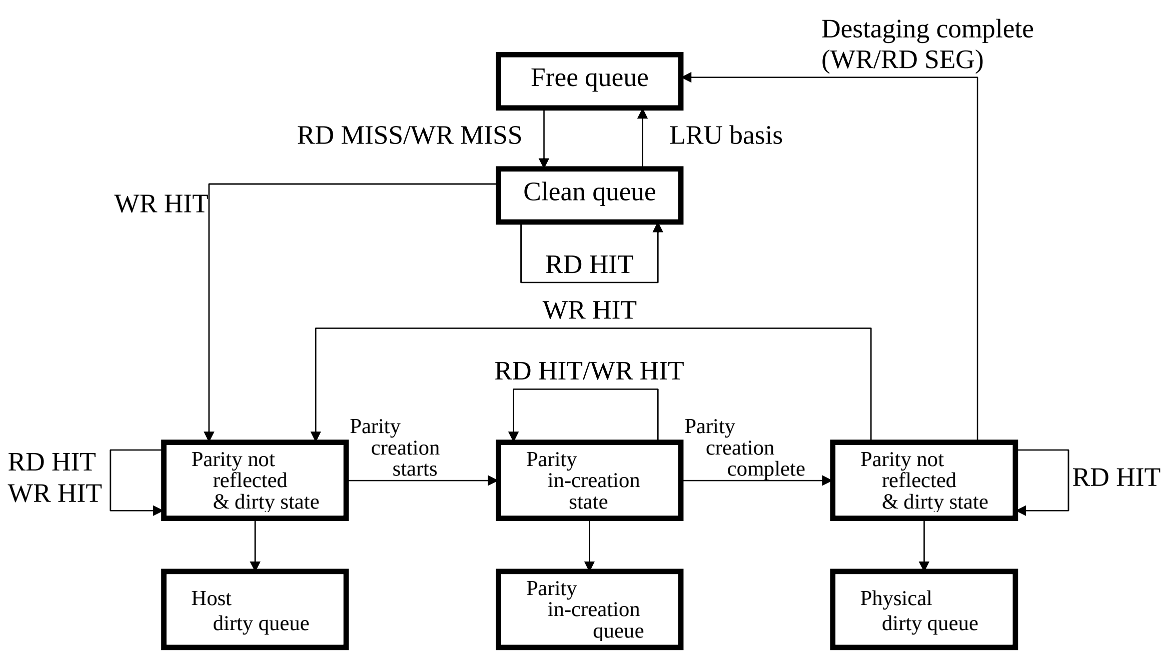
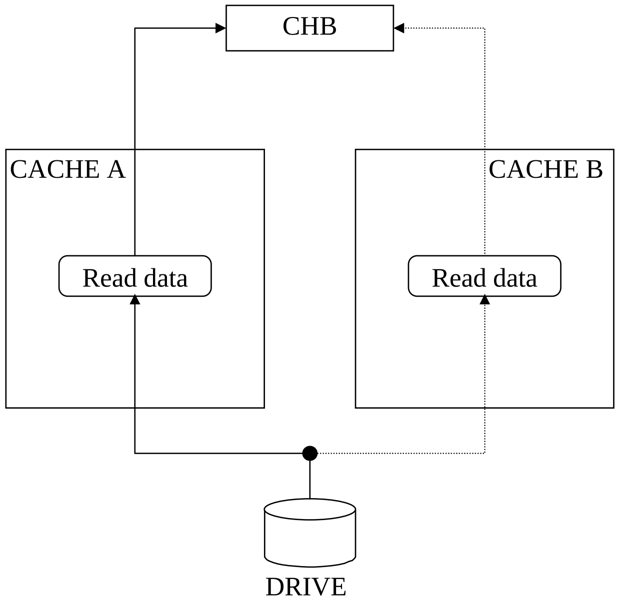


Fig. 3.3-3 Queue Segment State Transition Diagram

Copyright © 2012, Hitachi, Ltd.

(5) Cache usage in the read mode



The cache area to be used for destaging read data is determined depending on whether the result of evaluating the following expression is odd or even:

(CYL# x 15 + HD#) / 16

The read data is destaged into area A if the result is even and into area B if the result is odd.

Fig. 3.3-4 Cache Usage in the Read Mode

Read data is not duplexed and its destaging cache area is determined by the formula shown in Fig. 3.3-4. Staging is performed not only on the segments containing the pertinent block but also on the subsequent segments up to the end of track (for increased hit ratio). Consequently, one track equivalence of data is prefetched starting at the target block. This formula is introduced so that the cache activity ratios for areas A and B are even. The staged cache area is called the cache area and the other area NVS area.

Copyright © 2012, Hitachi, Ltd.

(6) Cache usage in the write mode

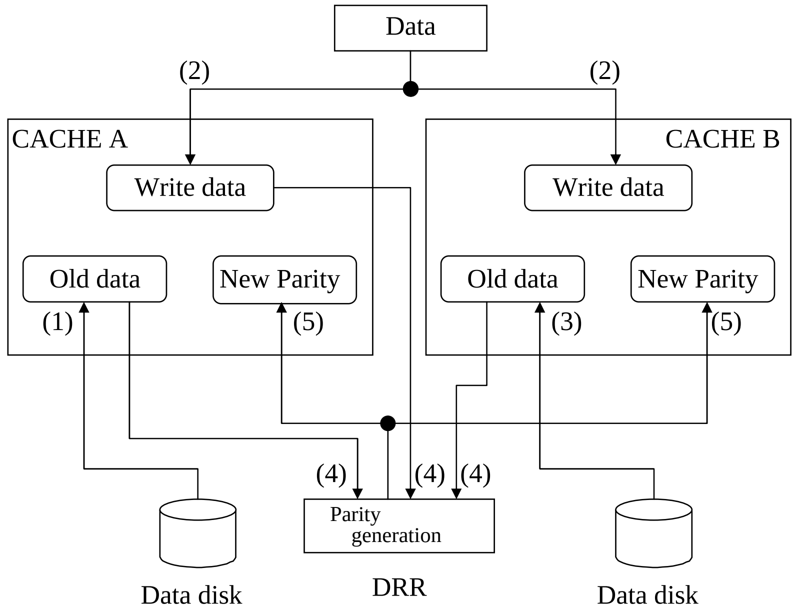


Fig. 3.3-5 Cache Usage in the Write Mode

This system handles write data (new data) and read data (old data) in separate segments as shown in Fig. 3.3-5 (not overwritten as in the conventional systems), whereby compensating for the write penalty.

(1) If the write data in question causes a cache miss, the data from the block containing the target record up to the end of the track is staged into a read data slot.
(2) In parallel with step (1), the write data is transferred when the block in question is established in the read data slot.
(3) The parity data for the block in question is checked for a hit or miss condition and, if a cache miss condition is detected, the old parity is staged into a read parity slot.
(4) When all data necessary for generating new parity is established, it is transferred to the DRR circuit in the DKB.
(5) When the new parity is completed, the DRR transfers it into the write parity slots for cache A and cache B (the new parity is handled in the same manner as the write data).

The reason for writing the write data into both cache areas is that data will be lost if a cache error occurs when it is not yet written on the disk.

Although two cache areas are used as explained above, the read data (including parity) is staged into either cache A or cache B simply by duplexing only the write data (including parity) (in the same manner as in the read mode).

 Copyright © 2012, Hitachi, Ltd.

(7) CFW-inhibited write-operation (with Cache single-side error)
The non RAID-type Disk systems write data directly onto disk storage in the form of cache through, without performing a DFW, when a cache error occurs. In this system, cache must always be passed, which fact disables the through operation. Consequently, the write data is duplexed, and a CFW-inhibited write-operation is performed; that is, when one cache storage system goes down, the end of processing status is not reported until the data write in the other cache storage system is completed. This process is called CFW-inhibited write-operation.

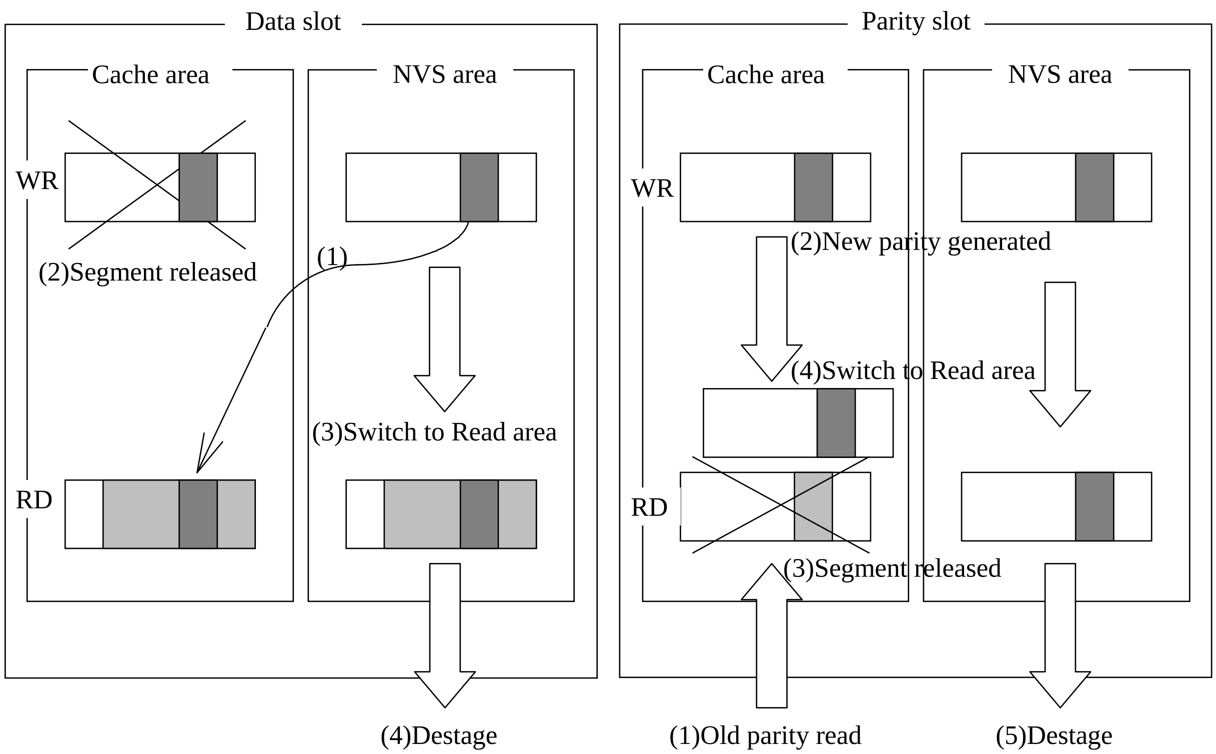
The control information necessary for controlling cache is stored in the shared memory.

Copyright © 2012, Hitachi, Ltd.

3.4 Destaging Operations

(1) Cache management in the destage mode (RAID5)

Destaging onto a drive is deferred until parity generation is completed. Data and parity slot transitions in the destage mode occur as shown in Fig. 3.4-1.



① The write data is copied from the NVS area into the read area.

② The write segment in the cache area is released.

③ Simultaneously, the segment in the NVS area is switched from write to read segment.

④ Destaging.

⑤ The read segment in the NVS area is released.

① Parity generation correction read (old parity) occurs.

② New parity is generated.

③ The old parity in the read segment is released.

④ The segments in the cache and NVS areas are switched from write to read segment.

⑤ Destaging.

⑥ The read segment in the NVS area is released.

Fig. 3.4-1 Cache Operation in the Destage Mode

Write data is stored in write segments before parity is generated but stored in read segments after parity is generated. When drive data is stored, therefore, the data from the read segment is transferred.

Copyright © 2012, Hitachi, Ltd.

(2) Cache management in the destage mode (RAID1)
Data slot is destaged to primary/secondary drive.

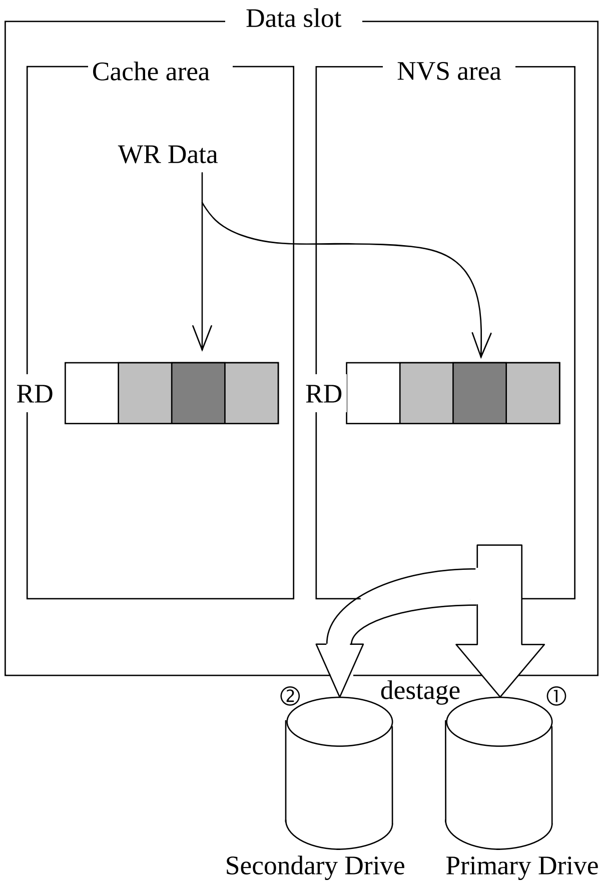


Fig. 3.4-2 RAID1 asynchronous destage

① Destage to primary drive.
② Destage to secondary drive.
③ The data read segment in the NVS area is released.

Copyright © 2012, Hitachi, Ltd.

(3) Blocked data write

The purpose of blocked data write is to reduce the number of accesses to the drive during destaging, whereby increasing the storage system performance. There are three modes of blocked data write: single-stripe blocking, multiple-stripe blocking, and drive blocking. These modes are briefly explained below.

- Single-stripe blocking

Two or more dirty segments in a stripe are combined into a single dirty data block. Contiguous dirty blocks are placed in a single area. If an unloaded block exists between dirty blocks, the system destages the dirty blocks separately at the unloaded block. If a clean block exists between dirty blocks, the system destages the blocks including the clean block.

- Multiple-stripe blocking

The sequence of stripes in a parity group are blocked to reduce the number of write penalties. This mode is useful for sequential data transfer.

- Drive blocking

In the drive blocking mode, blocks to be destaged are written in a block with a single drive command if they are contiguous when viewed from a physical drive to shorten the drive's latency time.

The single- and multiple-stripe blocking modes are also called in-cache blocking modes. The DMP determines which mode to use. The drive blocking mode is identified by the DSP.

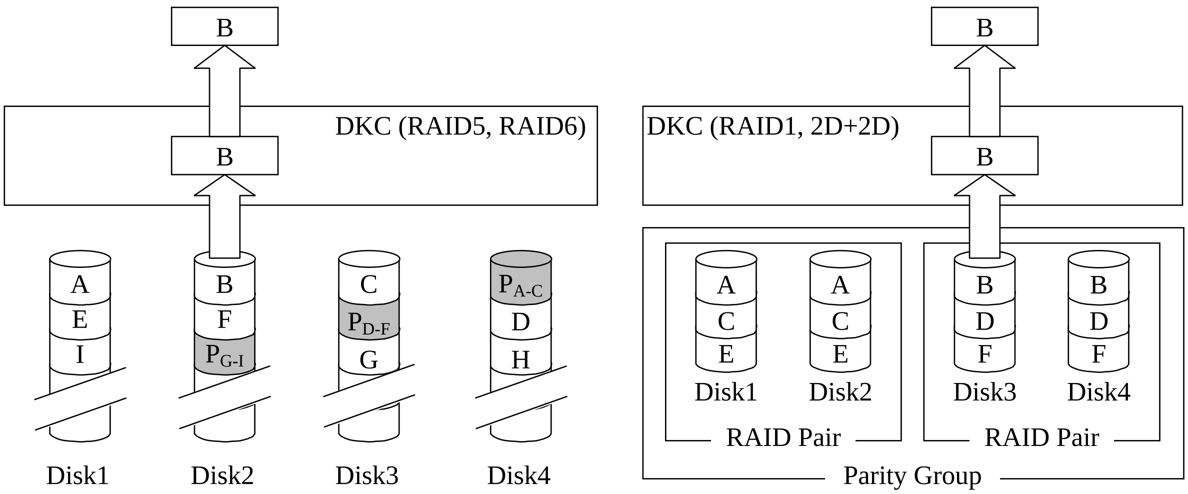
 Copyright © 2012, Hitachi, Ltd.

3.5 Operations Performed when Drive Errors Occur

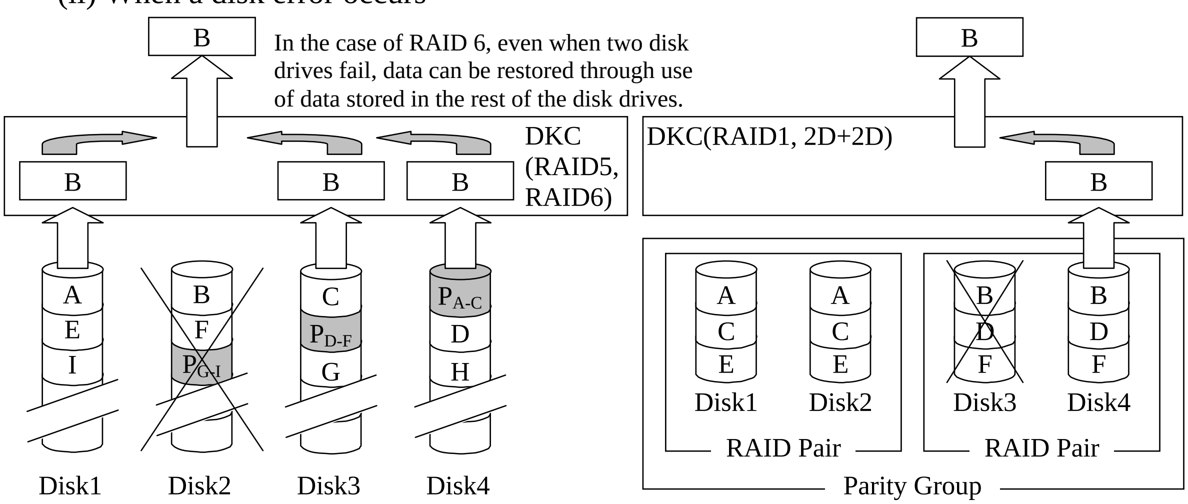
(1) I/O operations performed when drive errors occur
This system can recover target data using parity data and data stored on normal disk storage even when it cannot read data due to errors occurring on physical drives. This feature ensures non-disruptive processing of applications in case of drive errors. This system can also continue processing for the same reason in case errors occur on physical drives while processing write requests.

Fig. 3.5-1 shows the outline of data read processing in case a drive error occurs.

Request for reading data B
(i) Normal time



(ii) When a disk error occurs



A, B, C • • • :Data (A=A’, B=B’, C=C’)
P :Parity data

Fig. 3.5-1 Outline of Data Read Processing

Copyright © 2012, 2013, Hitachi, Ltd.

(2) Data integrity feature and drive errors

This system uses spare disk drives and reconfigures any drives that are blocked due to errors or drives whose error count exceeds a specified limit value using spare disks. (Drives belonging to a Parity Group with no LDEVs defined are not reconfigured.)

Since this processing is executed on the host in the background, this system can continue to accept I/O requests. The data saved on spare disks are copied into the original location after the error drives are replaced with new ones. But when the copy back mode is set to “No Copy Back”, and when copying to the same capacity spare disk, the copy back is not performed.

1. Dynamic sparing

This system keeps track of the number of errors that occurred, for each drive, when it executes normal read or write processing. If the number of errors occurring on a certain drive exceeds a predetermined value, this system considers that the drive is likely to cause unrecoverable errors and automatically copies data from that drive to a spare disk. This function is called dynamic sparing. In RAID1 method, this system is same as RAID5 dynamic sparing.

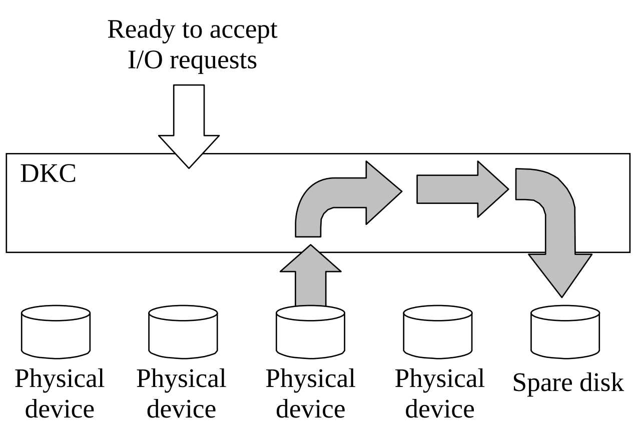


Fig. 3.5-2 Outline of the Dynamic Sparing Function

Copyright © 2012, Hitachi, Ltd.

2. Correction copy

When this system cannot read or write data from or to a drive due to an error occurring on that drive, it regenerates the original data for that drive using data from the other drives and the parity data, and copies it onto a spare disk. In RAID1 method, this system copies data from the another drive to a spare disk.

In the case of RAID 6, the correction copy can be made to up to two disk drives in a parity group.

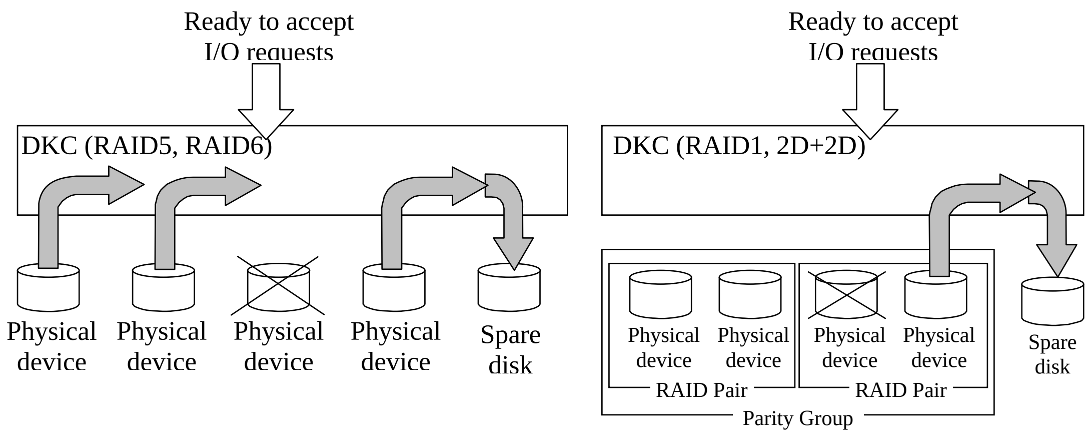


Fig. 3.5-3 Outline of the Correction Copy Function

3. Allowable number of copying operations

Table 3.5-1 Allowable number of copying operations

RAID level	Allowable number of copying operations
RAID1	Either the dynamic sparing or correction copy can be executed within a RAID pair.
RAID5	Either the dynamic sparing or correction copy can be executed within a parity group.
RAID6	The dynamic sparing and/or correction copy can be executed up to a total of twice within a parity group.

 Copyright © 2012, 2013, Hitachi, Ltd.

3.6 Inter Mix of Drives

3.6.1 Drives to be Connected

The models of disk units which are connectable with the DW700 storage system and the specifications of each disk unit are shown in Table 2.3-1 (THEORY02-190).

The DW700 storage system can connect up to 1152 disk drives mentioned above, though the number of connectable disk drives varies with the RAID configuration. These will be explained in detail in Section 3.6.2.

SVP displays each drive model as the following table.

Disk drive model	SVP screen	Drive Form Factor
DKS5C-K300SS	DKS5C-K300SS	2.5 inch SAS HDD
DKR5C-J600SS	DKR5C-J600SS	2.5 inch SAS HDD
DKR5D-J600SS	DKR5D-J600SS	2.5 inch SAS HDD
DKS5D-J600SS	DKS5D-J600SS	2.5 inch SAS HDD
DKS5E-J600SS	DKS5E-J600SS	2.5 inch SAS HDD
DKR5D-J900SS	DKR5D-J900SS	2.5 inch SAS HDD
DKS5D-J900SS	DKS5D-J900SS	2.5 inch SAS HDD
DKS5E-J900SS	DKS5E-J900SS	2.5 inch SAS HDD
DKR5E-J1R2SS	DKR5E-J1R2SS	2.5 inch SAS HDD
SLB5A-M200SS	SLB5A-M200SS	2.5 inch SSD
SLR5B-M200SS	SLR5B-M200SS	2.5 inch SSD
SLB5A-M400SS	SLB5A-M400SS	2.5 inch SSD
SLR5B-M400SS	SLR5B-M400SS	2.5 inch SSD
SLB5A-M800SS	SLB5A-M800SS	2.5 inch SSD
NFH1A-P1R6SS	NFH1A-P1R6SS	Flash Module Drive
DKR2D-H3R0SS	DKR2D-H3R0SS	3.5 inch SAS HDD
DKS2D-H3R0SS	DKS2D-H3R0SS	3.5 inch SAS HDD
DKR2E-H3R0SS	DKR2E-H3R0SS	3.5 inch SAS HDD
DKS2E-H3R0SS	DKS2E-H3R0SS	3.5 inch SAS HDD
NFH1B-P3R2SS	NFH1B-P3R2SS	Flash Module Drive
DKR2E-H4R0SS	DKR2E-H4R0SS	3.5 inch SAS HDD
DKS2E-H4R0SS	DKS2E-H4R0SS	3.5 inch SAS HDD

Copyright © 2012, 2013, Hitachi, Ltd.

Disk drive models (same capacity and same rotation speed) are intermixed in same ECC, Recommendation setting on SVP is following.

Disk drive model	Recommendation setting
DKS5x-K300SS	DKS5C-K300SS
DKR5x-J600SS	DKS5D-J600SS
DKS5x-J600SS	
DKR5x-J900SS	DKS5D-J900SS
DKS5x-J900SS	
DKR5x-J1R2SS	DKR5E-J1R2SS
SLB5x-M200SS	SLR5B-M200SS
SLR5x-M200SS	
SLB5x-M400SS	SLR5B-M400SS
SLR5x-M400SS	
SLB5x-M800SS	SLB5A-M800SS
NFH1x-P1R6SS	NFH1A-P1R6SS
DKR2x-H3R0SS	DKS2D-H3R0SS
DKS2x-H3R0SS	
NFH1x-P3R2SS	NFH1B-P3R2SS
DKR2x-H4R0SS	DKS2E-H4R0SS
DKS2x-H4R0SS	

 Copyright © 2012, 2013, Hitachi, Ltd.

When the HDD is replaced, the HDD which has the compatibility is following.

Before replacing	After replacing
DKS5x-K300SS	DKS5x-K300SS
DKR5x-J600SS	DKR5x-J600SS DKS5x-J600SS
DKS5x-J600SS	DKR5x-J600SS DKS5x-J600SS
DKR5x-J900SS	DKR5x-J900SS DKS5x-J900SS
DKS5x-J900SS	DKR5x-J900SS DKS5x-J900SS
DKR5x-J1R2SS	DKR5x-J1R2SS
SLB5x-M200SS	SLB5x-M200SS SLR5x-M200SS
SLR5x-M200SS	SLB5x-M200SS SLR5x-M200SS
SLB5x-M400SS	SLB5x-M400SS SLR5x-M400SS
SLR5x-M400SS	SLB5x-M400SS SLR5x-M400SS
SLB5x-M800SS	SLB5x-M800SS
NFH1x-P1R6SS	NFH1x-P1R6SS
DKR2x-H3R0SS	DKR2x-H3R0SS DKS2x-H3R0SS
DKS2x-H3R0SS	DKR2x-H3R0SS DKS2x-H3R0SS
NFH1x-P3R2SS	NFH1x-P3R2SS
DKR2x-H4R0SS	DKR2x-H4R0SS DKS2x-H4R0SS
DKS2x-H4R0SS	DKR2x-H4R0SS DKS2x-H4R0SS

x: A, B, C...

 Copyright © 2012, 2013, Hitachi, Ltd.

3.6.2 Specifications for coexistence of elements

Table 3.6.2-1 shows permitted coexistence of RAID levels and HDD types respectively.

Table 3.6.2-1 Specifications for Coexistence of Elements

Item	Specification	Remarks
Coexistence of RAID levels	• RAID1/RAID5/RAID6 can exist in the system.	
Coexistence of numbers of HDDs composing ECC group (coexistence of configurations 7D+1P, 3D+1P, 2D+2D, 14D+2P and 6D+2P)	• The numbers of HDDs inside an ECC group are 4HDDs or 8HDDs and they are applicable to RAID5 (7D+1P, 3D+1P), RAID1 (2D+2D) and RAID6 (14D+2P, 6D+2P). • All HDD types support RAID5 (7D+1P, 3D+1P), RAID1 (2D+2D) and RAID6 (14D+2P, 6D+2P) configuration. • 7D+1P, 3D+1P, 2D+2D, 14D+2P and 6D+2P can coexist in the system.	
Coexistence of HDD types	• HDD types can coexist in each ECC group. • The specification for selecting the spare HDD can be common in following cases. (1) More than the same capacity. (2) In the case of HDD, the same rotation speed. In the case of SSD, the same memory method. In the case of FMD, only FMD is available. (3) The same interface.	

Copyright © 2012, Hitachi, Ltd.

3.7 LDEV Formatting

3.7.1 High-Speed Format

3.7.1.1 Outlines

DKC can LDEV-Format to two or more ECC at the same time by providing the HDD with the LDEV formatting function.

Table 3.7.1.1-1

Item No.	Item	Contents
1	SVP operation	The operation is performed by selecting functions from the Maintenance menu.
2	Display of execution status	Display of the execution progress in the SVP message box (%)
3	Execution result	Normal/abnormal LDEV: Same indications as the conventional ones are displayed. Normal/abnormal PDEV: STATUS is displayed.
4	Recovery action when a failure occurs	Same as the conventional one. However, a retry is to be executed in units of ECC. (Because the LDEV-FMT is terminated abnormally in units of ECC when a failure occurs in the HDD.)
5	Operation of the SVP which is a high-speed LDEV-FMT object	When an LDEV-FMT of more than one ECC is instructed, the high-speed processing is performed.
6	PS/OFF or powering off	The LDEV formatting is suspended. No automatic restart is executed.
7	SVP PC powering off during execution of an LDEV-FMT	After the SVP is rebooted, the indication before the PC powering off is displayed in succession.
8	Execution of a high-speed LEDV-FMT in the status that the spare is saved	ECC of HDD which the spare is saved fails the high-speed LEDV-FMT, and changes to a low-speed format. (Because the low-speed format is executed after the high-speed format is completed, the format time becomes long.) After the high-speed LEDV-FMT is completed, execute the copy back of HDD which the spare is saved from SIM log and restore it.

Copyright © 2012, 2013, Hitachi, Ltd.

3.7.1.2 Estimation of LDEV Formatting Time

(1) DKxxx-JxxxSS/KxxxSS/HxxxSS

(a) High speed LDEV formatting

The format time of DKxxx-JxxxSS/KxxxSS/HxxxSS doesn’t depend on number of ECC, and be decided by capacity and the rotational speed of HDD.
It is an aim to the last in the standard time required, and the real format time may be different by RAID GROUP and a drive type.
Formatting time is indicated as follows.

Table 3.7.1.2-1

HDD Capacity/rotation Speed	Formatting Time (*3)	Time Out Value (*1)
4TB/7.2krpm	approx. 600min	910min
3TB/7.2krpm	approx. 530min	730min
1.2TB/10krpm	approx. 170min	270min
900GB/10krpm	approx. 150min	220min
600GB/10krpm	approx. 100min	150min
300GB/15krpm	approx. 45min	65min

(b) Slow LDEV formatting

The format time of DKxxx-JxxxSS/KxxxSS/HxxxSS is indicated as follows.
LDEV formatting is performed by slow LDEV formatting only.
Rough formatting time per 1TB/1PG without host I/O is indicated as follows (*2).

Table 3.7.1.2-2 15krpm

RAID Level		Formatting Time (*3)
RAID1	2D+2D	125min
RAID5	3D+1P	85min
	7D+1P	35min
RAID6	6D+2P	45min
	14D+2P	20min

Other capacity becomes time proportional to 1TB.

Table 3.7.1.2-3 10krpm

RAID Level		Formatting Time (*3)
RAID1	2D+2D	175min
RAID5	3D+1P	115min
	7D+1P	50min
RAID6	6D+2P	60min
	14D+2P	25min

Other capacity becomes time proportional to 1TB.

Copyright © 2012, 2013, Hitachi, Ltd.

Table 3.7.1.2-4 7.2krpm

RAID Level		Formatting Time (*3)
RAID1	2D+2D	240min
RAID5	3D+1P	160min
	7D+1P	70min
RAID6	6D+2P	80min
	14D+2P	35min

Other capacity becomes time proportional to 1TB.

(2) SSD

SSD doesn't have the self LDEV formatting function.
LDEV fomatting is performed by slow LDEV formatting only.
Rough formatting time per 1TB/1PG without host I/O is indicated as follows (*2).

Table 3.7.1.2-5

RAID Level		Formatting Time (*3)
RAID1	2D+2D	50min
RAID5	3D+1P	35min
	7D+1P	15min
RAID6	6D+2P	20min
	14D+2P	10min

The Formatting time becomes the same in 16 SSDs because the transmission of the format data doesn't arrive even at the limit of passing.

(3) Flash Module Drive

(a) High speed LDEV formatting

The format time of NFxxx-PxxxSS doesn't depend on number of ECC, and be decided by capacity and the rotational speed of HDD.
It is an aim to the last in the standard time required, and the real format time may be different by RAID GROUP and a drive type.
Formatting time is indicated as follows.

Table 3.7.1.2-6

HDD Capacity/rotation Speed	Formatting Time (*3)	Time Out Value (*1)
1.6TB	60min	100min
3.2TB	120min	190min

Copyright © 2012, 2013, Hitachi, Ltd.

(b) Slow LDEV formatting
The format time of NFxxx-PxxxSS is indicated as follows.
LDEV formatting is performed by slow LDEV formatting only.
Rough formatting time per 1TB/1PG without host I/O is indicated as follows (*2).

Table 3.7.1.2-7

RAID Level		Formatting Time (*3)
RAID1	2D+2D	15min
RAID5	3D+1P	15min
	7D+1P	10min
RAID6	6D+2P	10min
	14D+2P	5min

*1: The progress rate on SVP is displays as “99%” during the “Formatting Time” and the “Time Out Value”.
Because HDD executes the formatting, and the progress rate to the total capacity is not understood, the ratio at the elapsed time from the format beginning to the Formatting time required is displayed.

*2: If there is an I/O operation, the minimum formatting time is over 6 times as long as the discrete value, depending on the I/O load.

*3: The format time varies according to the generation of the drive in standard time distance.

NOTE: If the HDD types and configurations mentioned in above (1), (2) coexist, the format time required matches the one of the HDD type whose standard time required is the longest. As a result, the time required to start using the logical volumes is longer than the case of adding HDD one by one.
Therefore, when adding HDDs in the cases of above (1), (2), it is recommended to start the operation one by one from the HDD type whose standard time required is the shortest.

Copyright © 2012, Hitachi, Ltd.

3.7.2 Quick Format

3.7.2.1 Outlines

Quick Format provides the function to format in the background by making the volumes usable without waiting for the completion of the format when starting the format. The support specifications are shown below.

Table 3.7.2.1-1 Quick Format Specifications

Item No.	Item	Contents
1	Support HDD type	All HDD type support
2	Number of parity groups	The number of parity groups that Quick Format is possible at the same time is up to 36. The number of volumes is not limited if it is within 36 parity groups. In the case of four concatenations, the number of parity groups is four. In the case of two concatenations, the number of parity groups is two.
3	Combination with various P.P.	It is operable in combination with all P.P.
4	Execution opportunity	When performing a format from SVP or Web Console, you can select either Quick Format or the normal format.
5	Additional start in execution	Quick format can be executed additionally within a total of 36 parity groups including the parity group during execution during Quick Format execution.
6	Preparing Quick Format	When executing Quick Format, management information is created first. I/O access cannot be executed in the same way as the normal format in this period. Creating management information takes up to about one minute for one parity group, and up to about 36 minutes in case of 36 parity groups for the preparation.

(To be continued)

Copyright © 2012, Hitachi, Ltd.

(Continued from the preceding page)

Item No.	Item	Contents
7	Blocking and restoring the volume	• When the volume during Quick Format execution is blocked for maintenance, the status of the volume (during Quick Format execution) is stored in the storage system. When the volume is restored afterwards, the volume status becomes “Normal (Quick Format)”. • When all of the volumes during Quick Format in the parity group are blocked, the number of parity groups that are during Quick Format displayed in the ‘Logical Device’ window of Storage Navigator and ‘Maintenance’ window of SVP decreases equally to the number of the blocked parity groups. However, the number of parity groups that can additionally execute Quick Format will not increase. The number of parity groups that can additionally execute Quick Format can be calculated with the following calculating formula; 36-X-Y. (Legend) X: The number of parity groups during Quick Format execution displayed in the window. Y: The number of parity groups with all of the volumes in the parity group blocked during Quick Format execution.
8	Operation at the time of PS OFF/ON	After P/S ON, Quick Format restarts.
9	Restrictions	• Quick Format cannot be executed to the journal volume of Universal Replicator, external volume, and virtual volume. • Volume Migration and Quick Restore of ShadowImage cannot be executed to a volume during Quick Format. • The Prestaging of Cache Residency Manager cannot be executed to the volume during Quick Format.

Copyright © 2012, Hitachi, Ltd.

3.7.2.2 Data security of volumes during Quick Format

The Quick Format control table is kept on SM, but it is also stored in SSD in case SM volatilizes so that it recovers from SSD in the case of volatilization start.

 Copyright © 2012, 2013, Hitachi, Ltd.

3.7.2.3 Quick Format time

Quick Format executes the format in background while executing I/O from HOST. Therefore, the Quick Format time may change significantly depending on the number of I/O from HOST or other conditions.
The following table shows the Quick Format time without I/O. The Quick Format time with I/O may be twice or three times of the following time.

Table 3.7.2.3-1 Quick Format Time

Drive capacity	Format time
200GB	5h
300GB	8h
400GB	10h
600GB	15h
800GB	20h
900GB	22h
1.6TB	40h
3TB	75h
3.2TB	80h
4TB	100h

- The time above shows the time when Quick Format is executed to all areas of the parity group, and when Quick Format is executed to a part of LDEV in the parity group, the time will be faster in proportion to the capacity of LDEV.
- The Quick Format time with I/O may be over five times of the time above.
- When Quick Format is executed to multiple parity groups, the time becomes slower than the time above depending on the number of parity groups. The proportion of the time becoming slower is generally two times slower when the number of parity groups are 15, and three times slower when the number of parity groups are 30.
- The time above might be up to four times slower depending on the capacity of cache memories and the number of MPBs.
- When Quick Format is executed to parity groups with different drive capacities at the same time, calculate the time with the parity group of the largest capacity.
- When the RAID level is RAID1, the formatting time becomes about half compared with the above-mentioned time.

Copyright © 2012, Hitachi, Ltd.

3.7.2.4 Performance during Quick Format

Quick Format executes the format in background while executing I/O from HOST. Therefore, it may influence the HOST performance. The following table shows the proportion of the performance influence. (However, this is only a rough standard, and it may change depending on the conditions.)

Table 3.7.2.4-1 Performance during Quick Format

I/O types	Performance when the ratio shows 100% at normal condition
Random read	80%
Random write to the unformatted area	20%
Random write to the formatted area	60%
Sequential read	90%
Sequential write	90%

Copyright © 2012, Hitachi, Ltd.

3.7.2.5 Combination with other maintenance

Table 3.7.2.5-1 Combination with Other Maintenance

Item No.	Maintenance Operation	Operation during Quick Format
1	Drive copy/correction copy	The processing is possible as well as the normal volumes, but unformatted area is skipped.
2	Conventional format	The conventional format is executable for the volumes that Quick Format is not executed.
3	Volume maintenance blockade	It is possible to block the volumes instructed by SVP for the volumes during Quick Format.
4	Volume forcible restore	If forcible restore is executed after the maintenance blockade, it returns to Quick Formatting.
5	Verify consistency check	Possible. However, the Verify consistency check for the unformatted area is skipped.
6	PDEV replacement	Possible as usual
7	P/K replacement	Possible as usual

3.7.2.6 SIM when Quick Format finished

After Quick Format is finished, SIM = 0x410100 is output when executing Quick Format from SVP.
When all Quick Format is finished, the above SIM is output if Quick Format is executed from Storage Navigator while executing Quick Format from SVP.
When Quick Format is executed only from Storage Navigator, SIM is not output.

Copyright © 2012, Hitachi, Ltd.

3.8 Ownership Management

3.8.1 Confirmation and Definitions of requests and issues

Table 3.8.1-1 Confirmation and Definitions of requests and issues

#	Request	Case
1	Maximize system performance by using MPU effectively	Initial set up (Define Configuration & Install) Ownership management resources are Installed. At performance tuning
2	Troubleshoot in the case of problems related to ownership	Troubleshoot
3	Confirm resources allocated to each MPU	Ownership management resources are Installed. At performance tuning Troubleshoot
4	Maintain performance for resources allocated to specific MPU	Ownership management resources are Installed. Installation of MPU

3.8.1.1 Request #1

Request
Maximize system performance by using MPU effectively.

Issue
Way to distribute resources to balance load of each MPU.

How to realize
(1) User directly allocates resources to each MPU.
(2) User does not allocate resources. Resources are allocated to each MPU automatically.

Case
(A) Define Configuration & Install
Target resource: LDEV
Setting IF: SVP

(B) Ownership management resources are installed
Target resources: LDEV/External VOL/JNLG
Setting IF: SVP/Storage Navigator/CLI/RMLib

Copyright © 2012, Hitachi, Ltd.

3.8.1.2 Request #2

Request

Maximize system performance by using MPU effectively.

Issue

Way to move resources to balance load of each MPU.

How to realize

User directly requests to move resources.

Case

Performance tuning

Target resources:	LDEV/External VOL/JNLG
Setting IF:	Storage Navigator/CLI/RMLib

3.8.1.3 Request #3

Request

Troubleshoot in the case of problems related to ownership.

Issue

Way to move resources required for solving problems.

How to realize

Maintenance personnel directly requests to move resources.

Case

Troubleshoot

Target resources:	LDEV/External VOL/JNLG
Setting IF:	Storage Navigator/CLI/RMLib

Copyright © 2012, Hitachi, Ltd.

3.8.1.4 Request #4

Request
Confirm resources allocated to each MPU.

Issue
Way to reference resources allocated to each MPU.

How to realize
User directly request to reference resources.

Case

(A) Before ownership management resources are installed
Target resources: LDEV/External VOL/JNLG
Referring IF: Storage Navigator/CLI/Report (XPDT)/RMLib

(B) Performance tuning
Target resources: LDEV/External VOL/JNLG
Referring IF: Storage Navigator/CLI/Report (XPDT)/RMLib

(C) Troubleshoot
Target resources: LDEV/External VOL/JNLG
Referring IF: Storage Navigator/CLI/Report (XPDT)/RMLib

 Copyright © 2012, Hitachi, Ltd.

3.8.1.5 Request #5

Request

Maintain performance for resources allocated to specific MPU.

Issue

Way to move resources allocated to each MPU automatically and, way to prevent movement of resources during Installation of MPU.

How to realize

Resources are NOT allocated/ moved automatically to the MPU that user specified.

Case

(A) When installing ownership management resources, preventing allocation of resources to the Auto Assignment "Disable" MPU.

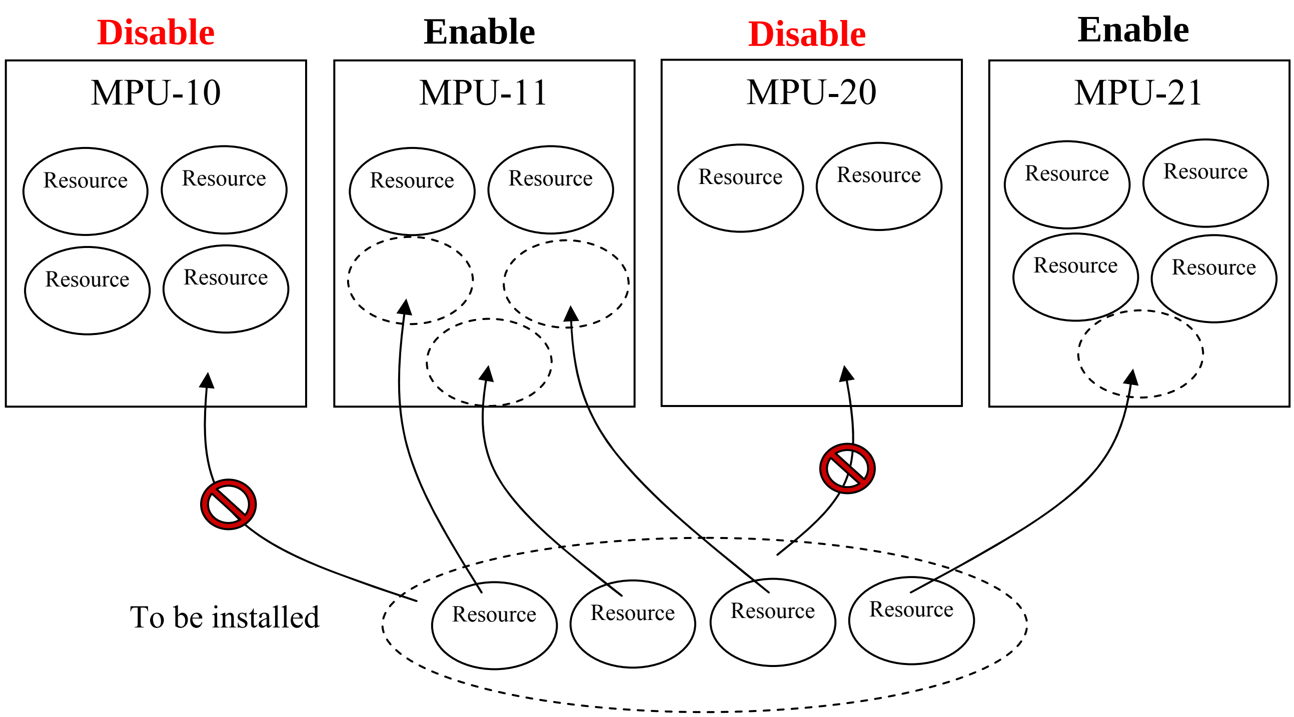


Fig. 3.8.1.5-1 Request #5 (A)

Copyright © 2012, Hitachi, Ltd.

3.8.1.6 Process flow

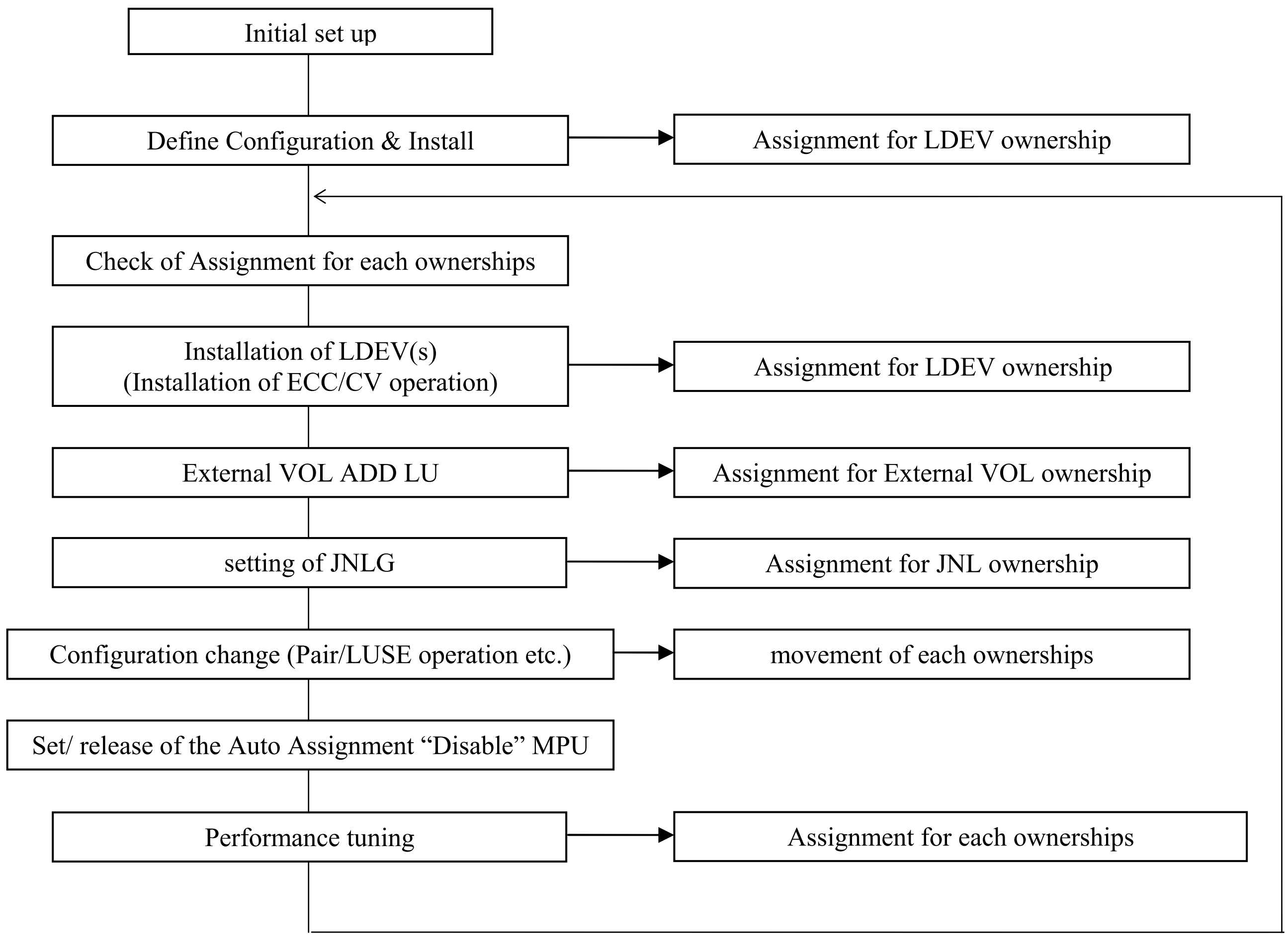


Fig. 3.8.1.6-1 Process flow

 Copyright © 2012, Hitachi, Ltd.

3.8.2 Resource allocation Policy

(1) Both User-specific and automation allocation based on common policy Allocate resources to each MPU equally.

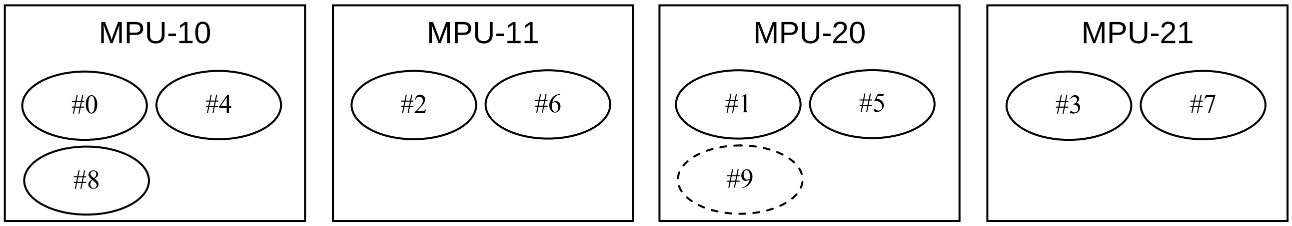


Fig. 3.8.2-1 Resource allocation Policy (1)

(2) Additionally, user-specific allocation can consider the weight of each device.

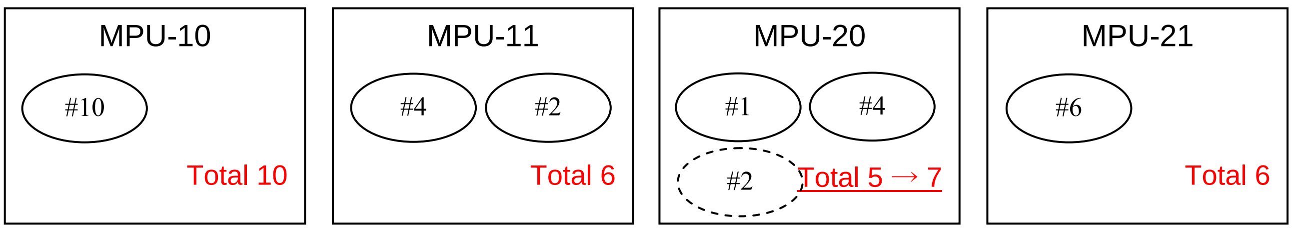


Fig. 3.8.2-2 Resource allocation Policy (2)

But, automation allocation cannot consider the weight of each device.

Copyright © 2012, 2013, Hitachi, Ltd.

3.8.2.1 Automation allocation

Allocate resources to each MPU equally independently in each type.

Table 3.8.2.1-1 Automation allocation

Owner type	Device type	Unit	Leveling
LDEV	SAS	ECC Gr.	LDEV num.
	SSD/FMD	LDEV	LDEV num.
	DP VOL	LDEV	LDEV num.
Ext. VOL	—	Ext. VOL	Ext. VOL num.
JNLG	—	JNLG	JNLG num.

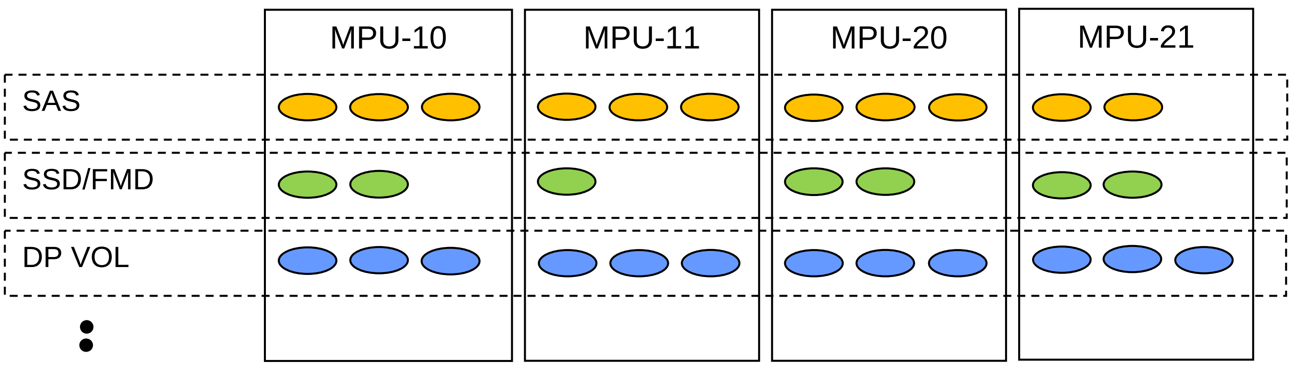


Fig. 3.8.2.1-1 Automation allocation

Copyright © 2012, 2013, Hitachi, Ltd.

3.8.2.1.1 Automation allocation (SAS)

Unit: ECC Gr.
Leveling: LDEV num.

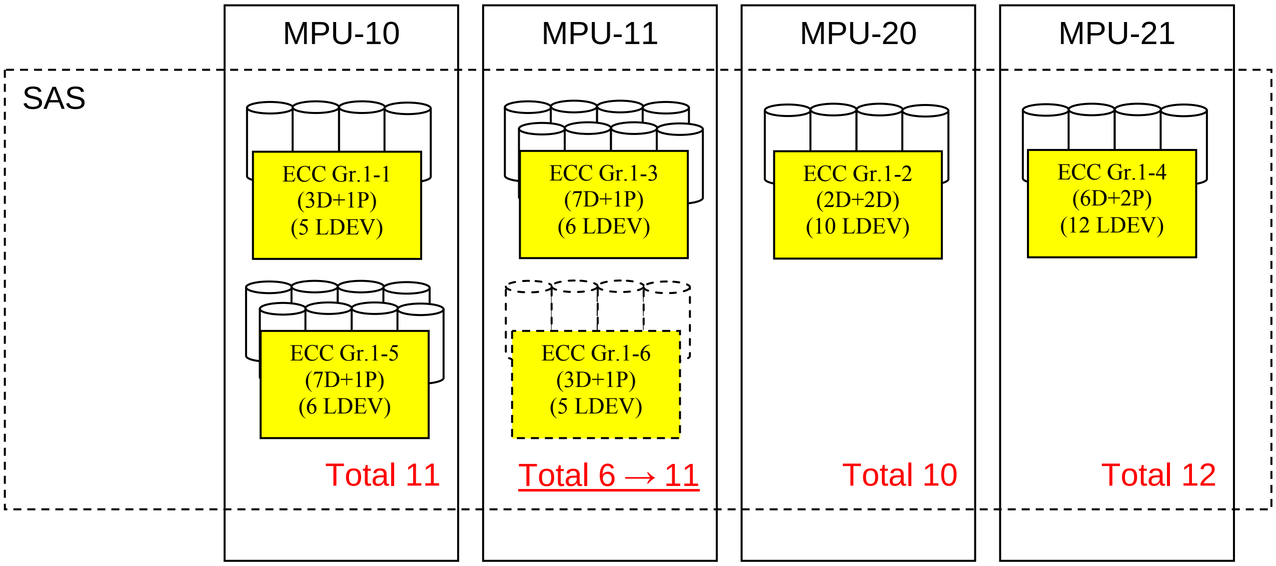


Fig. 3.8.2.1.1-1 Automation allocation (SAS)

3.8.2.1.2 Automation allocation (SSD, FMD/DP VOL)

Unit: ECC Gr.
Leveling: LDEV num.

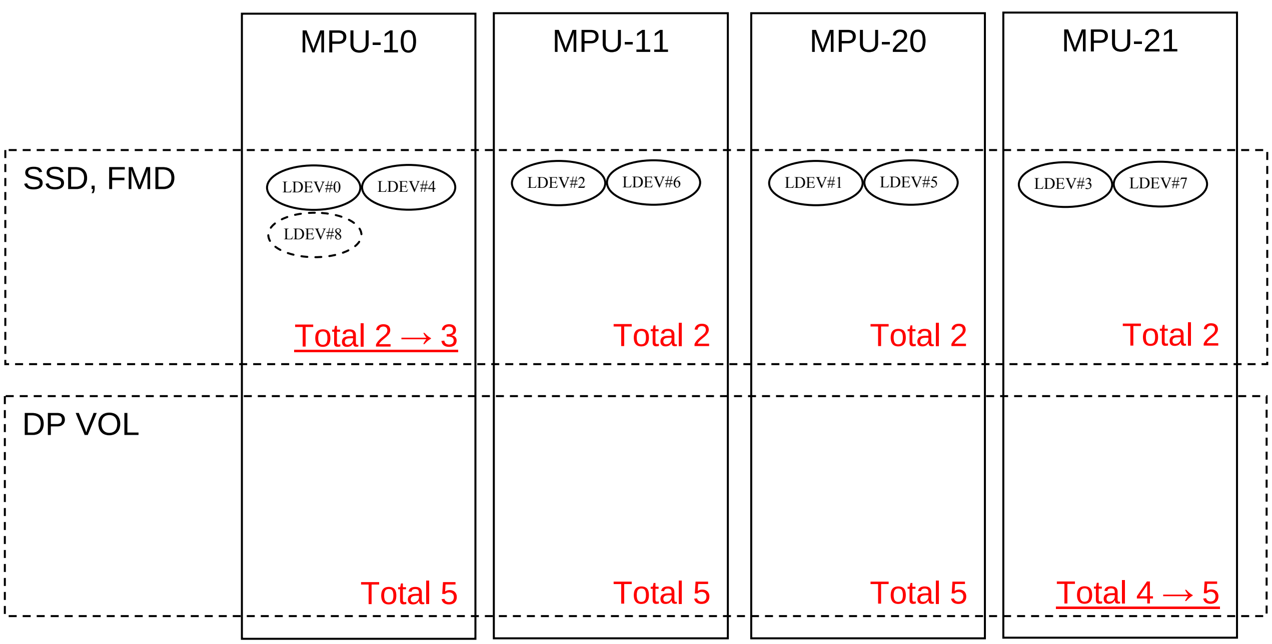


Fig. 3.8.2.1.2-1 Automation allocation (SSD, FMD/DP VOL)

 Copyright © 2012, Hitachi, Ltd.

3.8.2.1.3 Automation allocation (Ext. VOL)

Unit: Ext. VOL
Leveling: Ext. VOL num. (not Ext. LDEV num.)

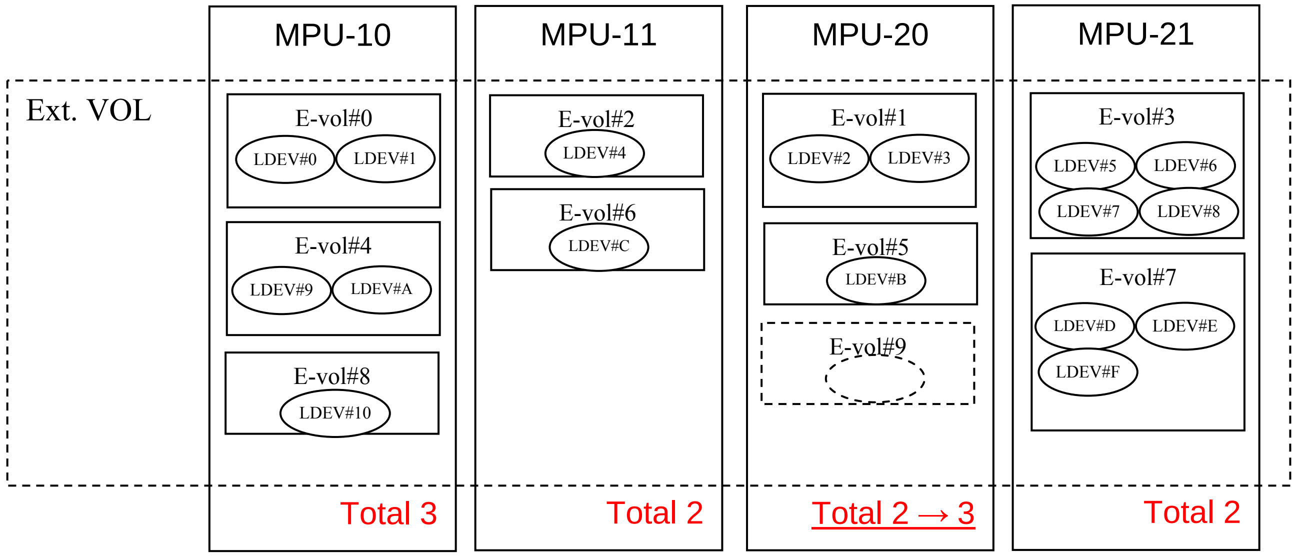


Fig. 3.8.2.1.3-1 Automation allocation (Ext. VOL)

3.8.2.1.4 Automation allocation (JNLG)

Unit: JNLG
Leveling: JNLG num. (not JNL VOL num.)

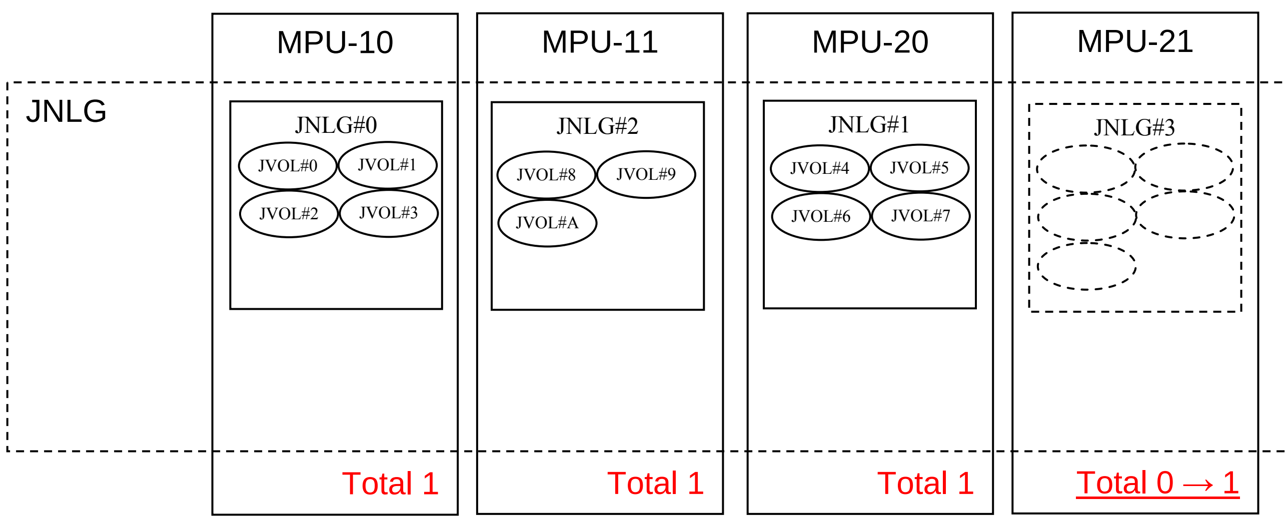


Fig. 3.8.2.1.4-1 Automation allocation (JNLG)

Copyright © 2012, Hitachi, Ltd.

3.8.3 MPU block

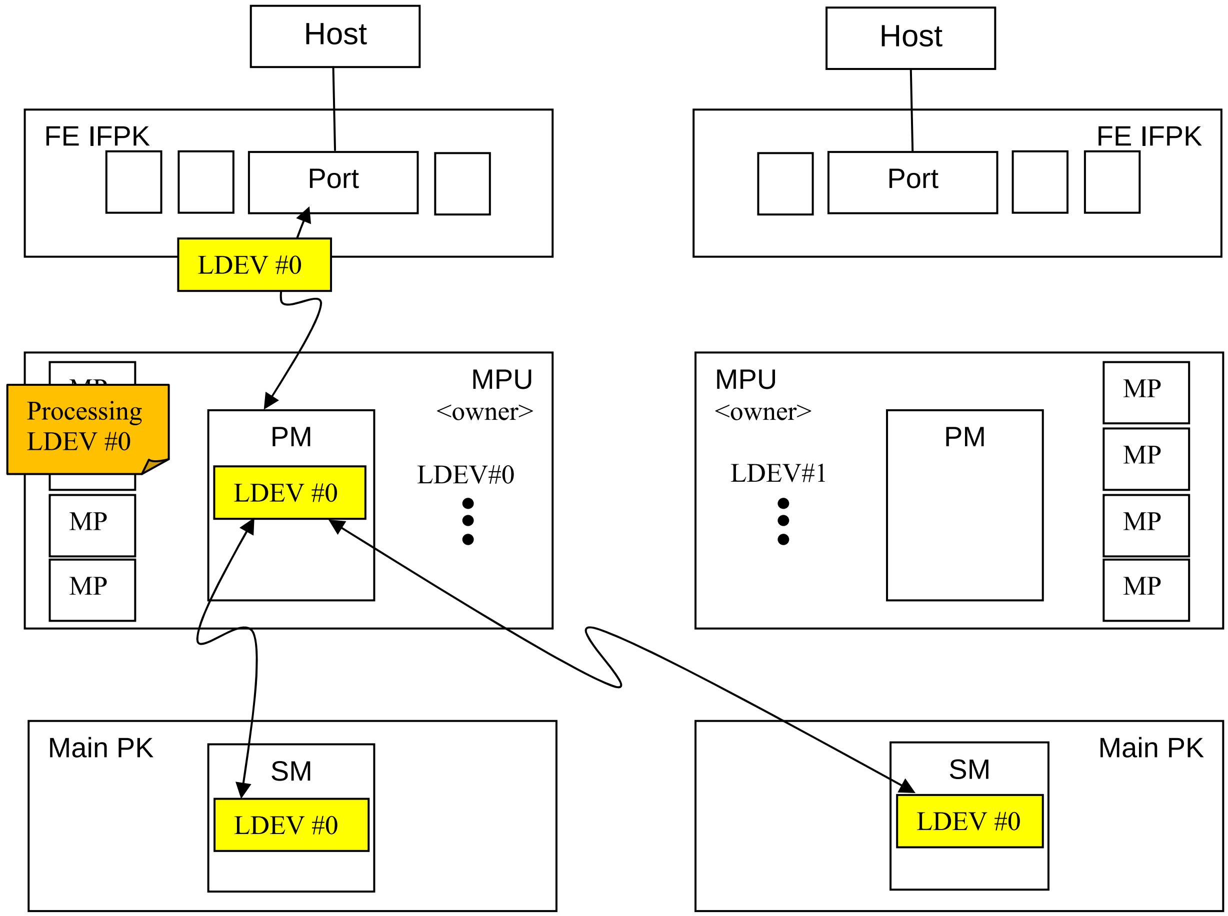


Fig. 3.8.3-1 MPU block

Copyright © 2012, Hitachi, Ltd.

3.8.3.1 MPU block for maintenance

Step1. Start moving the ownership.

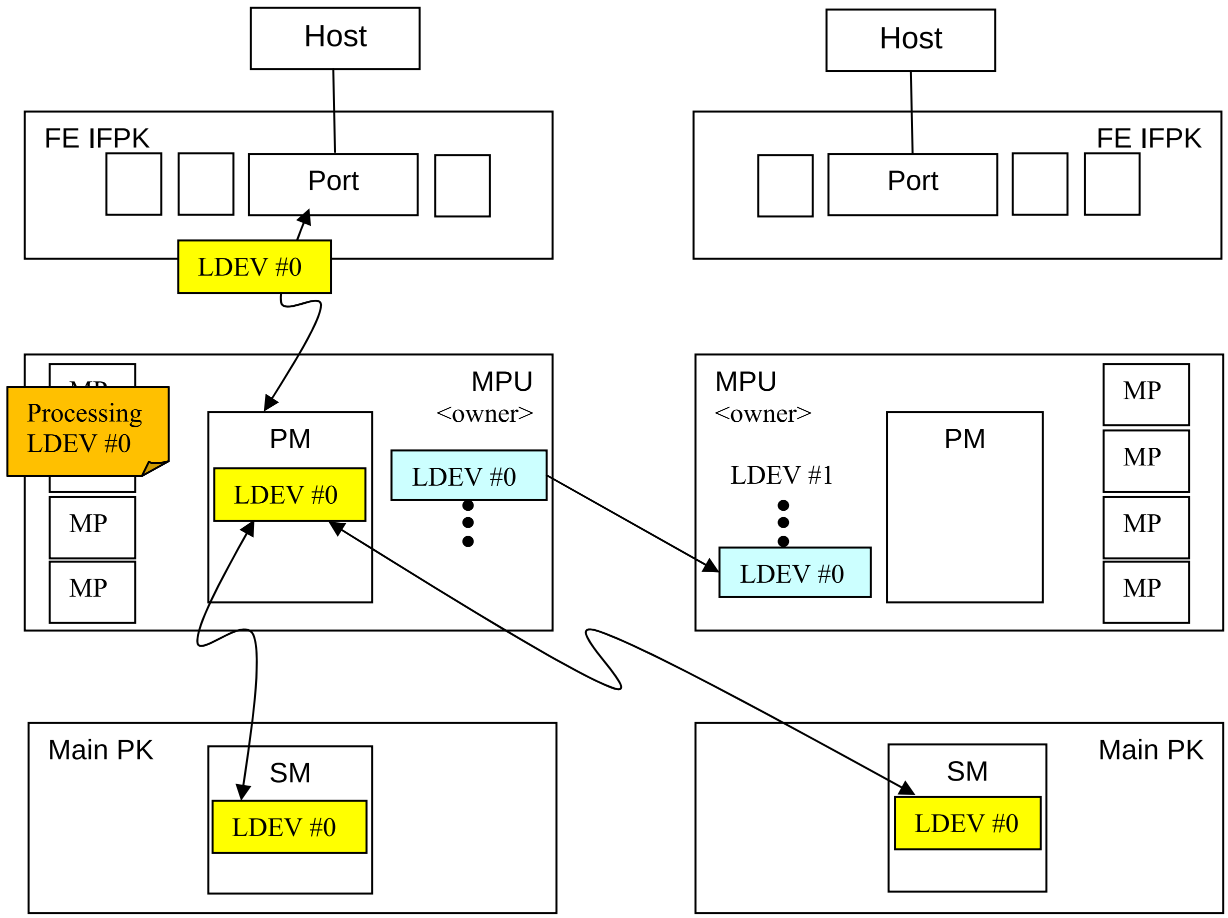


Fig. 3.8.3.1-1 MPU block for maintenance (1)

Copyright © 2012, Hitachi, Ltd.

Step2. Switch MPU, to which I/O is issued, to the target MPU (to which the ownership is moved).

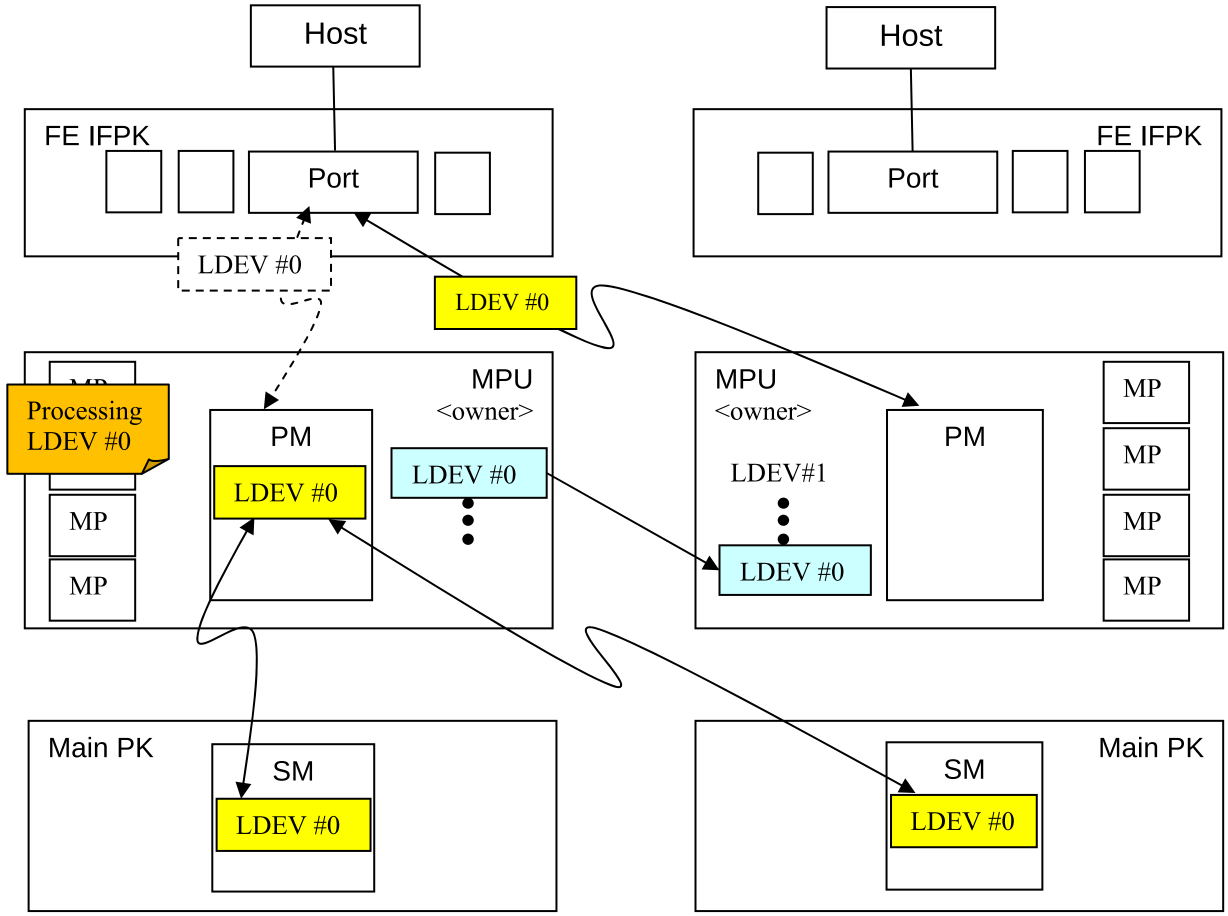


Fig. 3.8.3.1-2 MPU block for maintenance (2)

 Copyright © 2012, Hitachi, Ltd.

Step3. Complete the ongoing processing in the source MP whose ownership is moved. (New processing is not performed in the source MP.)

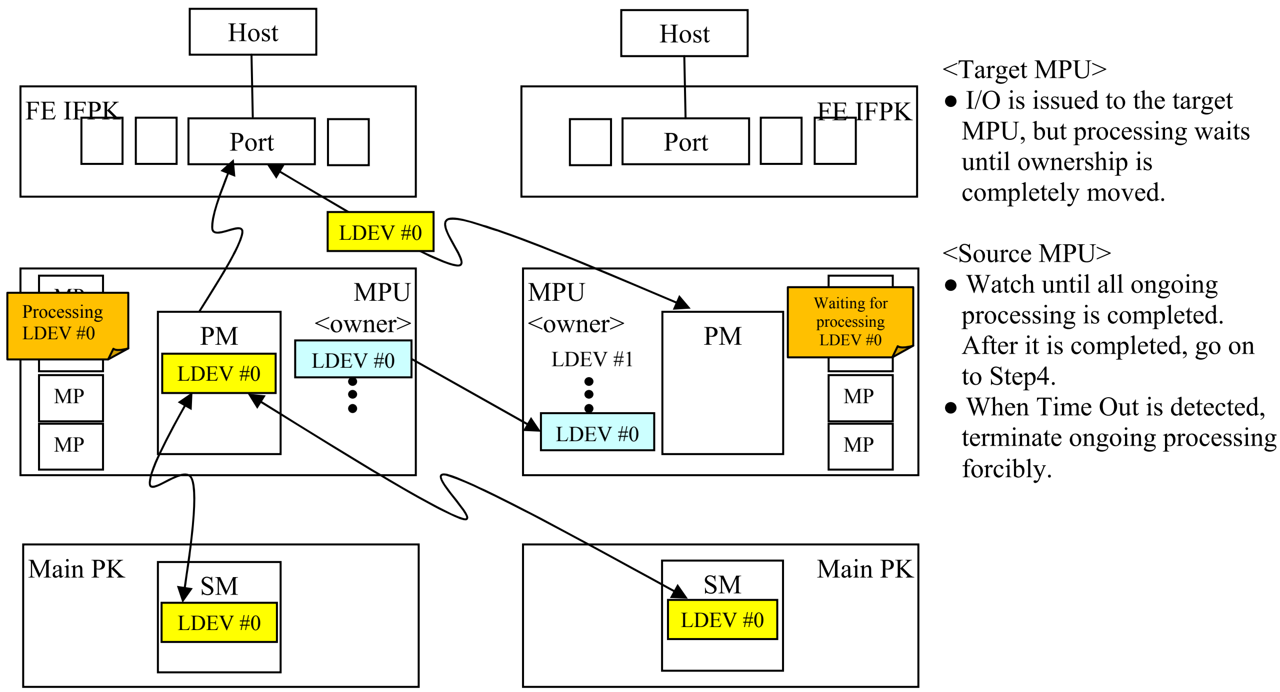


Fig. 3.8.3.1-3 MPU block for maintenance (3)

Step4. Disable PM information in the source MP whose ownership is moved.

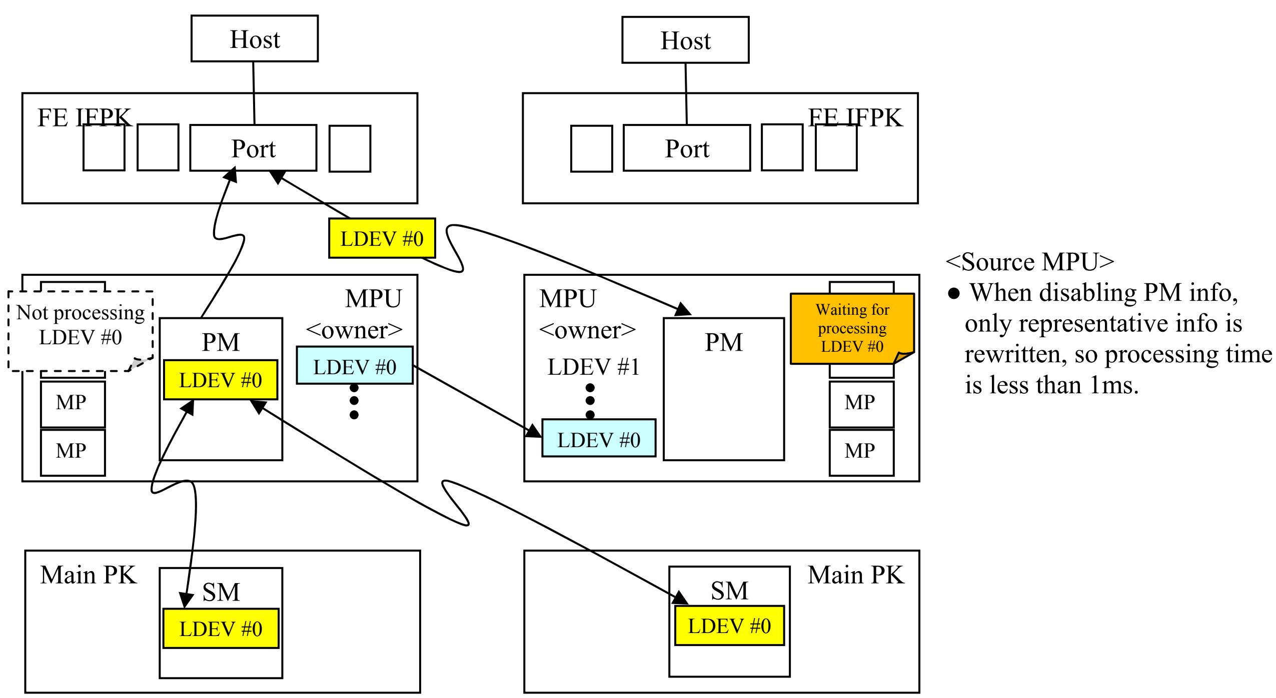
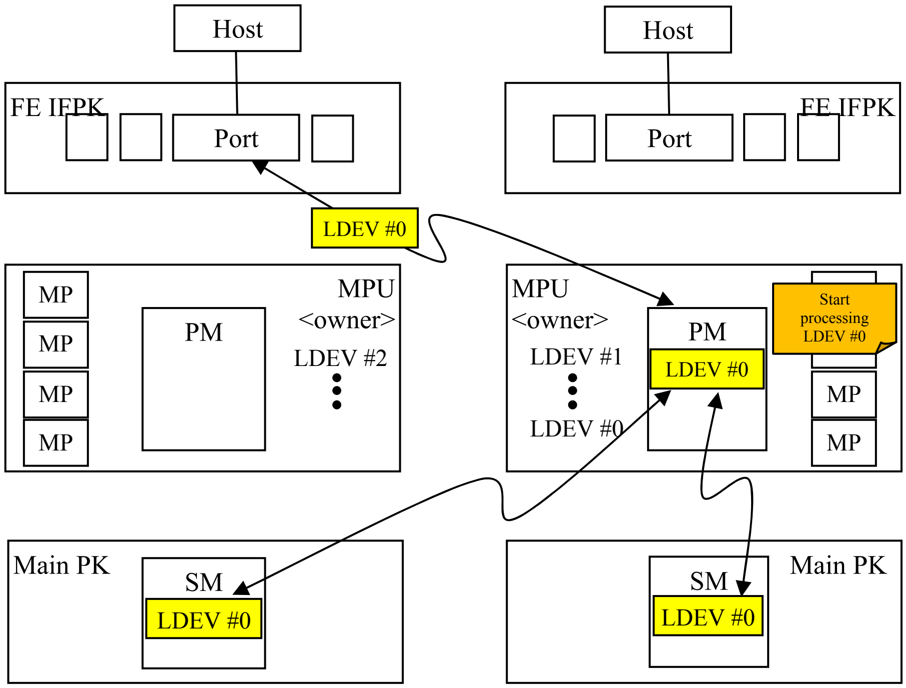


Fig. 3.8.3.1-4 MPU block for maintenance (4)

Step5. Moving ownership is completed, and the processing starts in the target MPU.

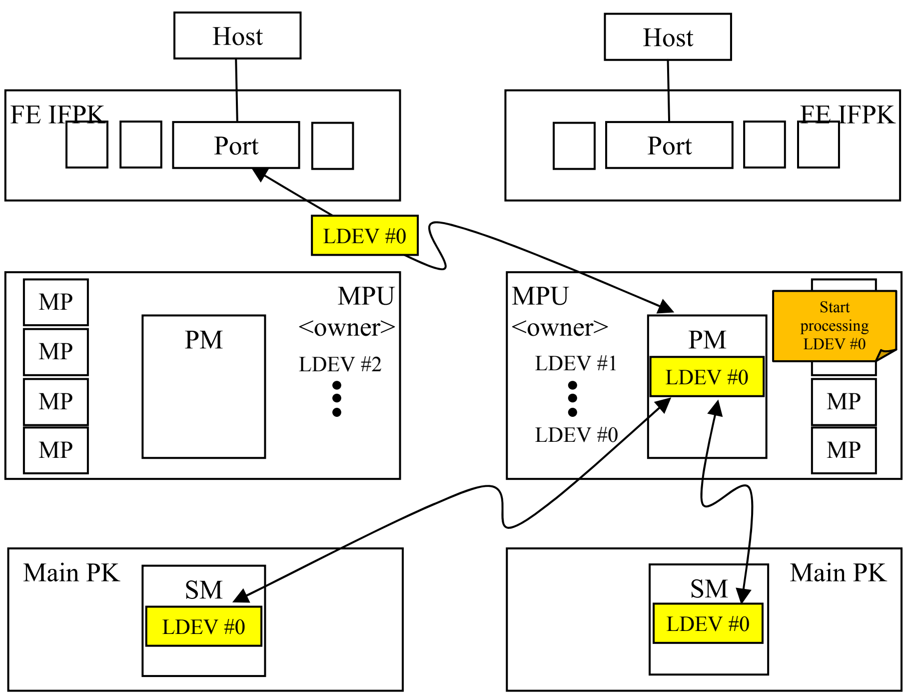


<Target MPU>

- Immediately after processing is started, SM is accessed, so access performance to control info is degraded compared to before moving ownership.
- As reading PM progresses, access performance to control info is improved.

Fig. 3.8.3.1-5 MPU block for maintenance (5)

Step6. Perform Step1. ~ Step5. for all resources under the MPU blocked and after they are completed, block MPU.



<Moving ownership>

- Move resources that are related to LUSE, ShadowImage, UR, and TI synchronously.
- If they are moved at a time, performance would be affected significantly, so move them in phase as long as maintenance time is permissible.

Fig. 3.8.3.1-6 MPU block for maintenance (6)

 Copyright © 2012, Hitachi, Ltd.

3.8.3.2 MPU blocked due to failure

Step1. Detect that all MPs in the MPU are blocked, and choose MPU that takes over the ownership.

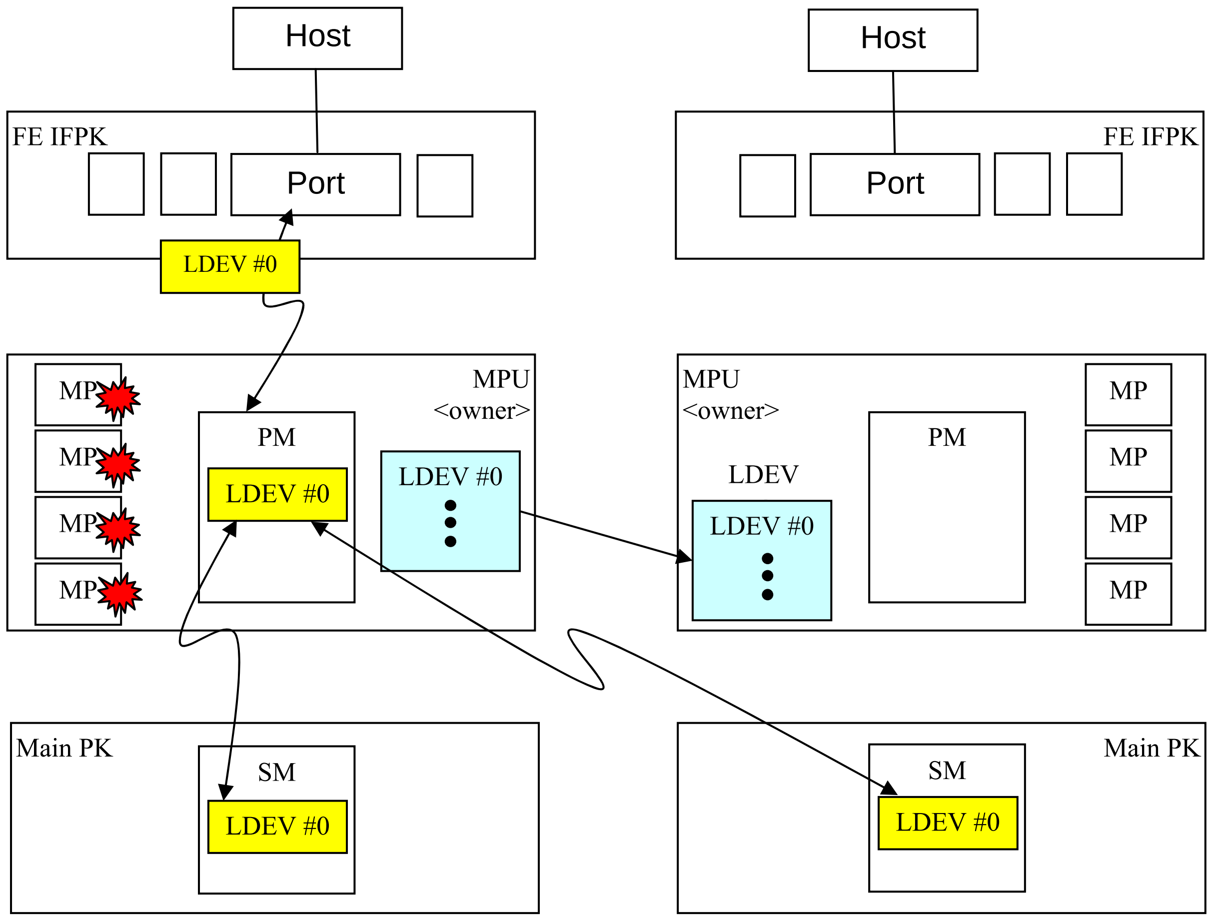


Fig. 3.8.3.2-1 MPU blocked due to failure (1)

 Copyright © 2012, Hitachi, Ltd.

Step2. Switch MPU, to which I/O is issued, to MPU that takes over the ownership.

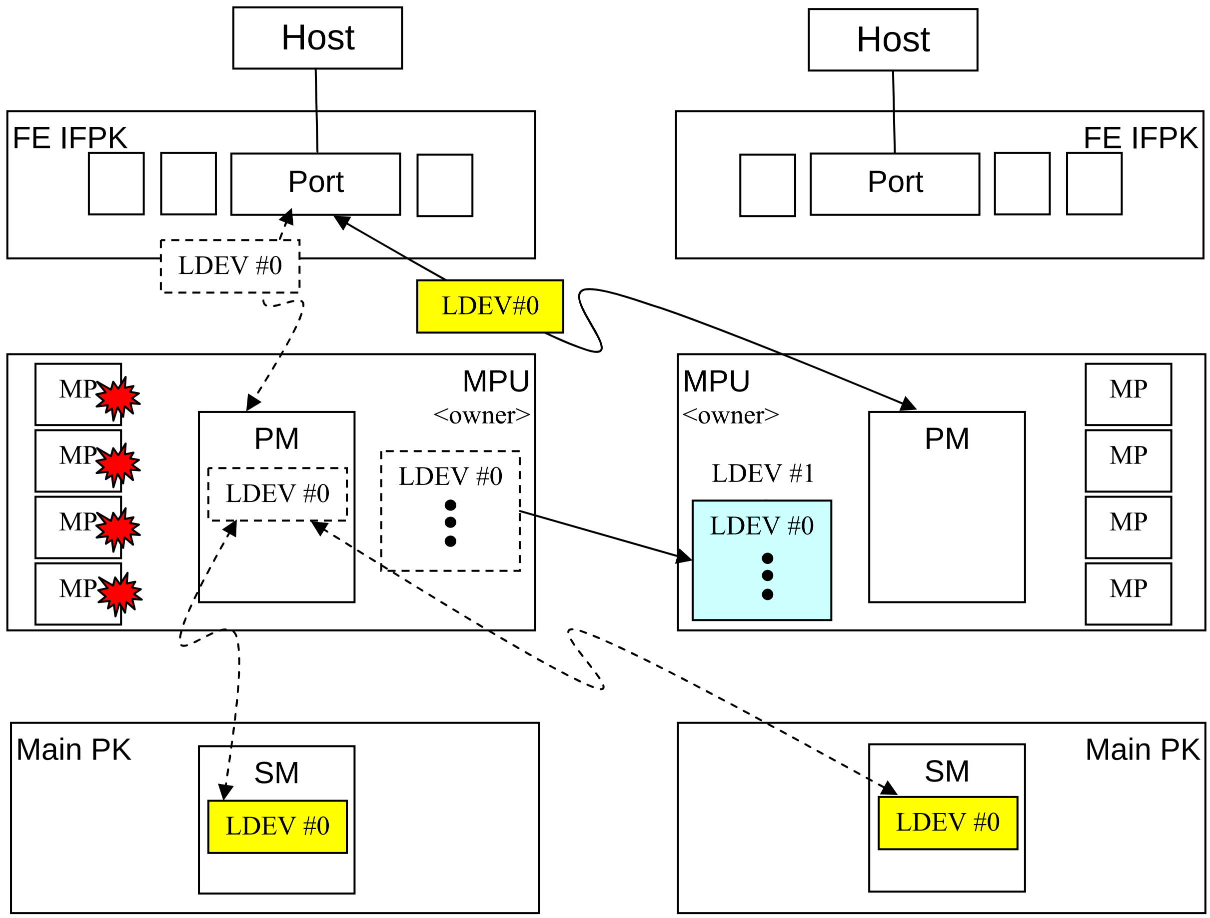
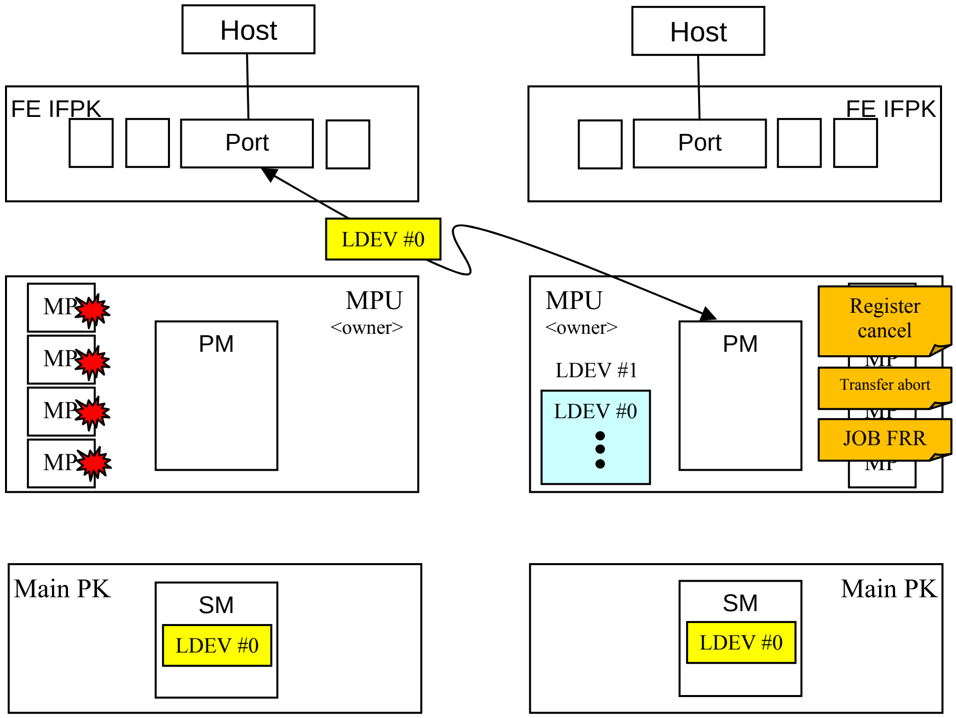


Fig. 3.8.3.2-2 MPU blocked due to failure (2)

 Copyright © 2012, Hitachi, Ltd.

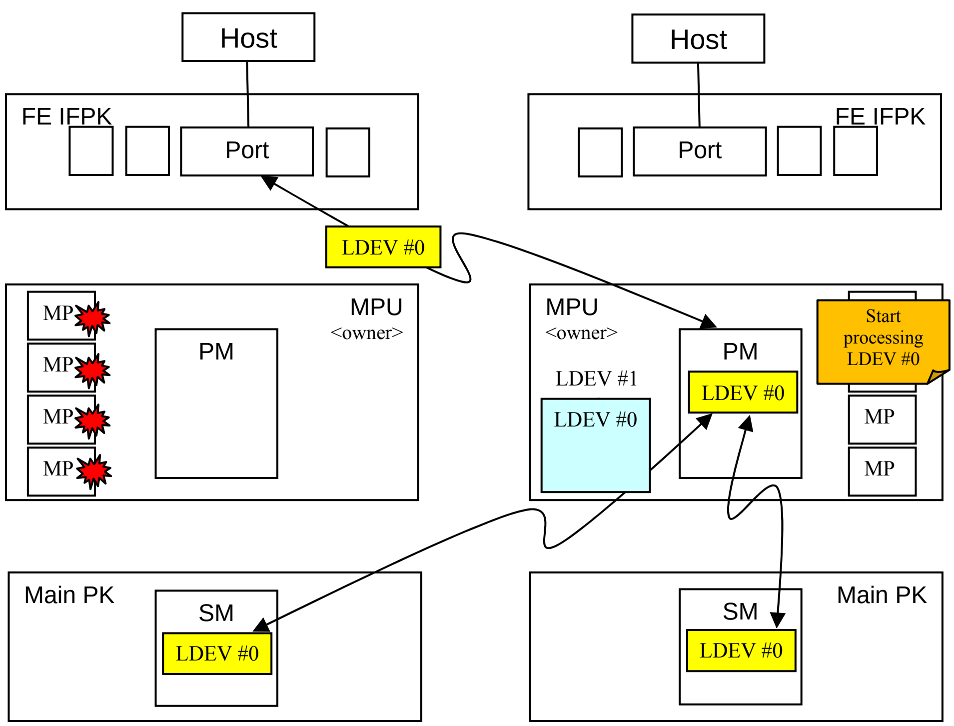
Step3. Perform WCHK1 processing at the initiative of MPU that takes over the ownership.



<WCHK1 processing>
- Issue a request for cancelling the request for processing received from WCHK1 MPU to all MPs.
- Issue an abort instruction of data transfer started from WCHK1 MPU.
- WCHK1 MPU performs post-processing of ongoing JOB.

Fig. 3.8.3.2-3 MPU blocked due to failure (3)

Step4. WCHK1 processing is completed, and the processing starts in the target MPU.



<Target MPU>
- Immediately after processing is started, SM is accessed, so access performance to control information is degraded compared to before moving ownership.
- As processes of importing information in PM progress, access performance to control info is improved.

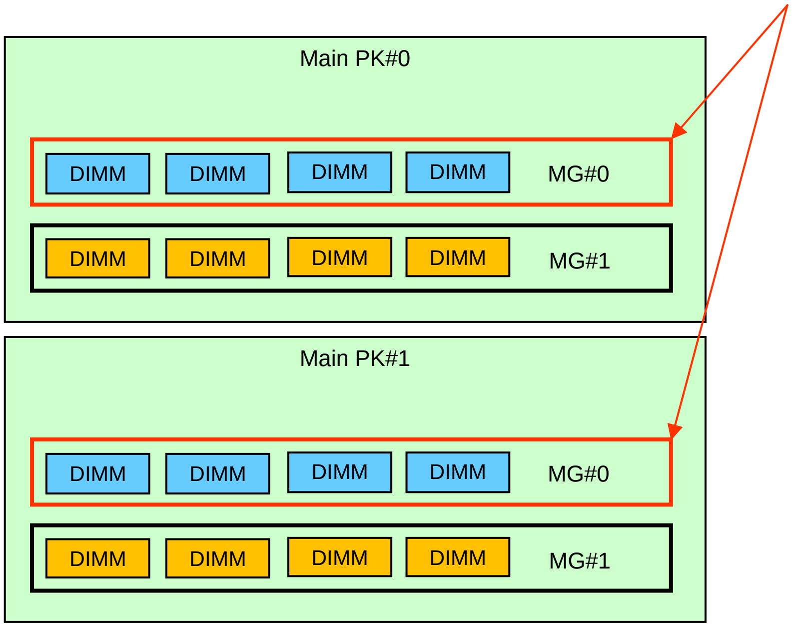
Fig. 3.8.3.2-4 MPU blocked due to failure (4)

Copyright © 2012, Hitachi, Ltd.

3.9 Cache Architecture

3.9.1 Physical installation of Main PK/DIMM

Installation unit DIMM 4x2=8



DIMM type

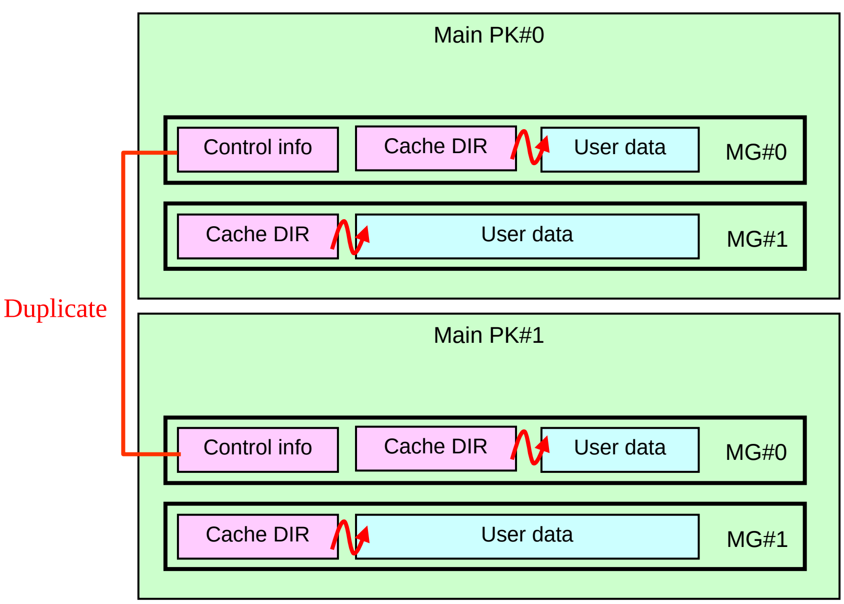
Type	GB/ DIMM	GB/MG	GB/DKC
C32G	4	16	32
C64G	8	32	64
C128G	16	64	128

Relationship b/w PK capacity and installed option

PK capacity	C32G	C64G	C128G
16GB	1	0	0
32GB	2	0	0
32GB	0	1	0
48GB	1	1	0
64GB	0	2	0
64GB	0	0	1
80GB	1	0	1
96GB	0	1	1
128GB	0	0	2

Fig. 3.9.1-1 Physical installation of Main PK/DIMM

3.9.2 Consolidated Cache DIR and user data in MG



<Memory allocation specification>
(1) fix control information to MG#0

(2) Consolidate Cache DIR and User data in MG

<Allocated capacity per MG>

(1) Control information
4GB ~ 20GB

(2) Cache DIR
Fixed to 0.5GB

(3) User data
MG capacity – (1) – (2)
3.5GB ~ 59.5GB (MG#0)
15.5GB ~ 63.5GB (MG#1)

Fig. 3.9.2-1 Consolidated Cache DIR and user data in MG

Copyright © 2012, Hitachi, Ltd.

3.9.3 Maintenance/Failure Blockade Specification

3.9.3.1 Blockade Unit

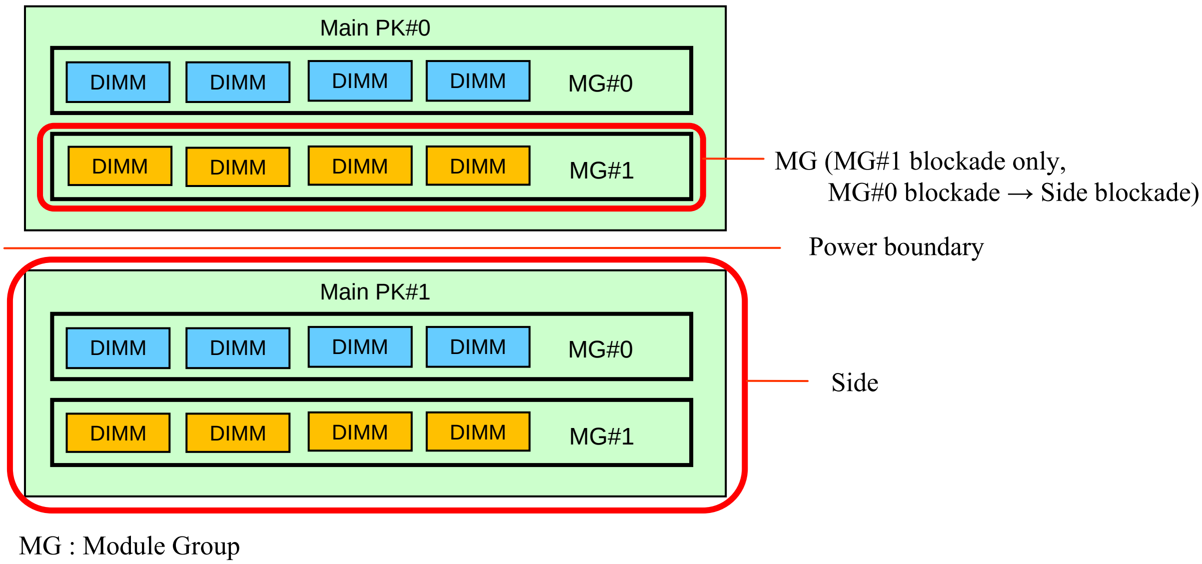


Fig. 3.9.3.1-1 Blockade Unit

 Copyright © 2012, Hitachi, Ltd.

3.9.4 Cache Control

3.9.4.1 Cache DIR PM read and PM/SM write

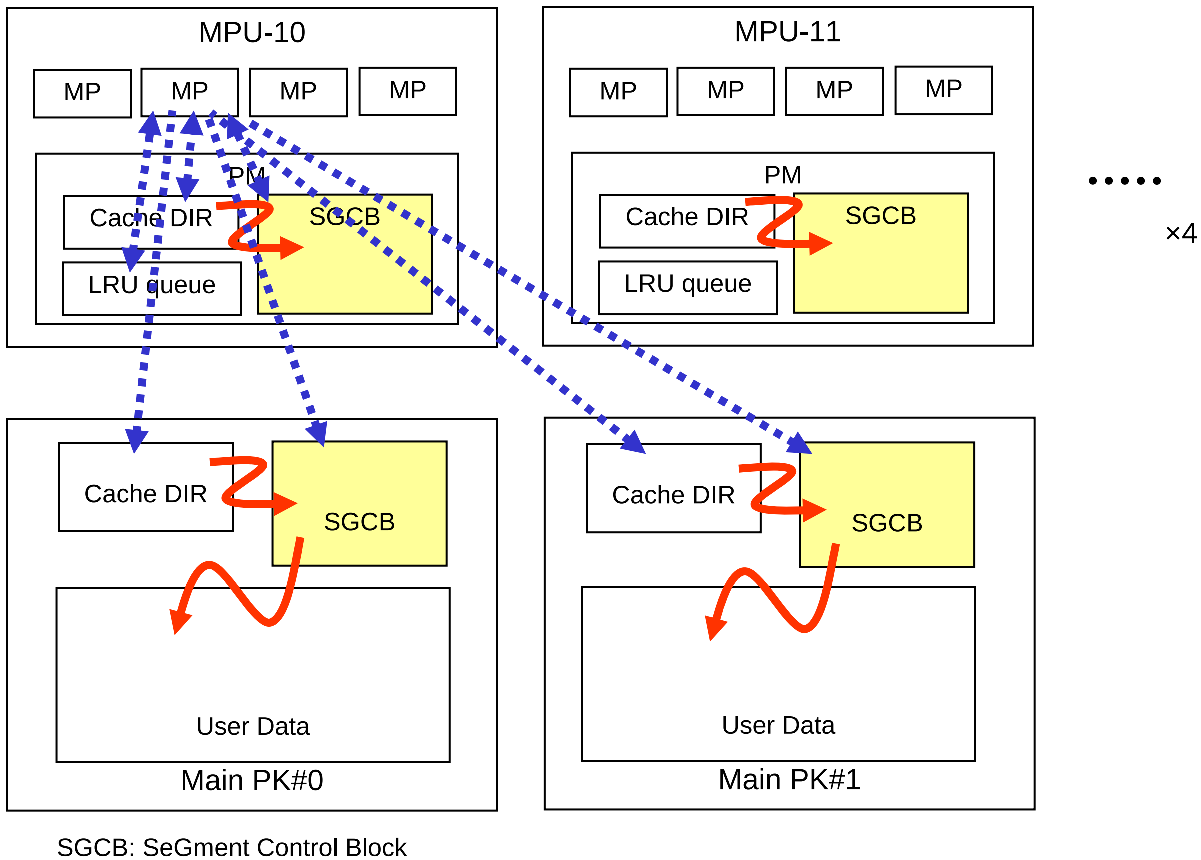


Fig. 3.9.4.1-1 Cache DIR PM read and PM/SM write

Copyright © 2012, Hitachi, Ltd.

3.9.4.2 Cache Segment Control Image

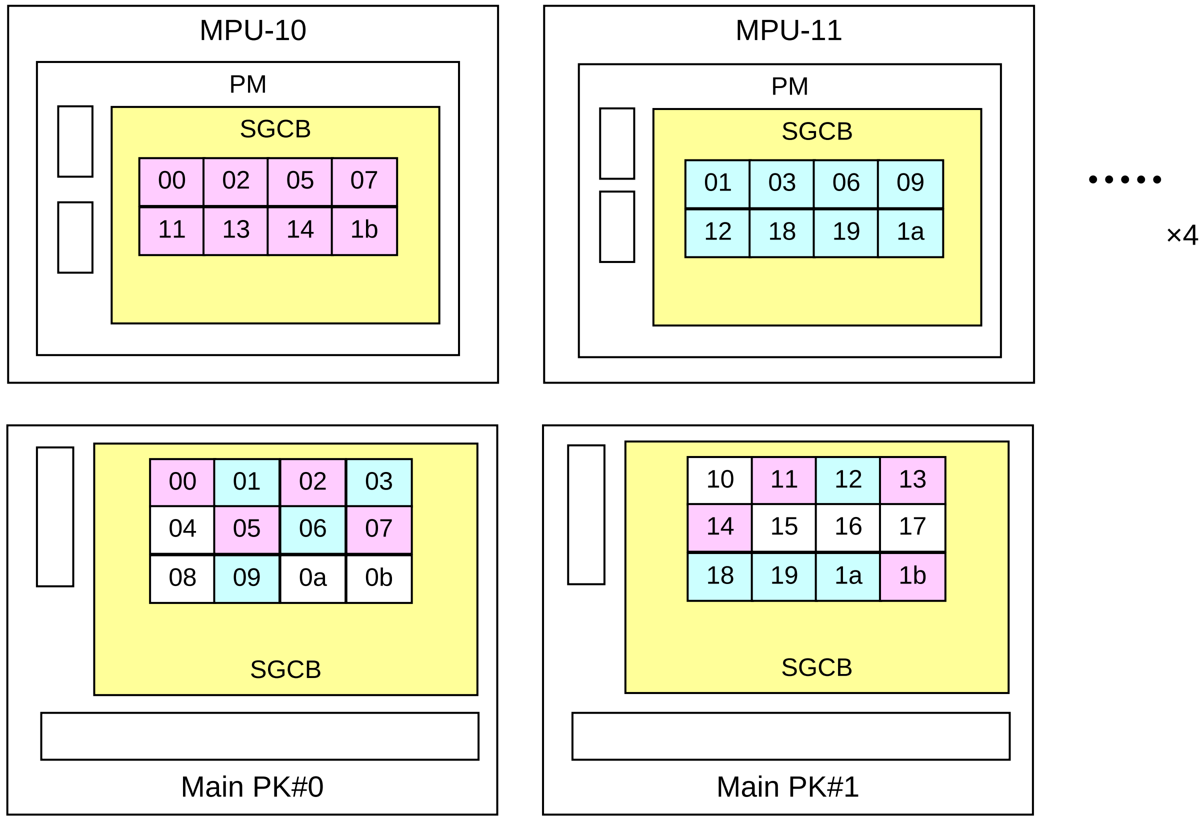


Fig. 3.9.4.2-1 Cache Segment Control Image

Copyright © 2012, Hitachi, Ltd.

3.9.4.3 Initial Setting (Cache Volatile)

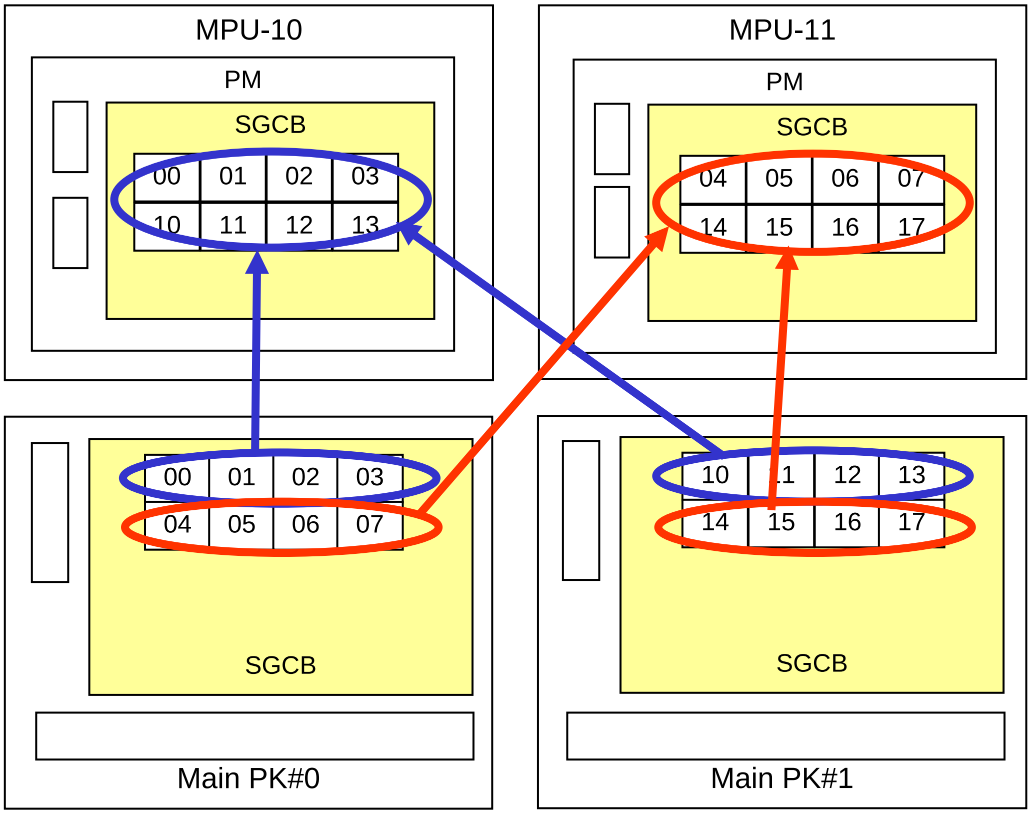


Fig. 3.9.4.3-1 Initial Setting (Cache Volatile)

 Copyright © 2012, Hitachi, Ltd.

3.9.4.4 Main Blade Replace

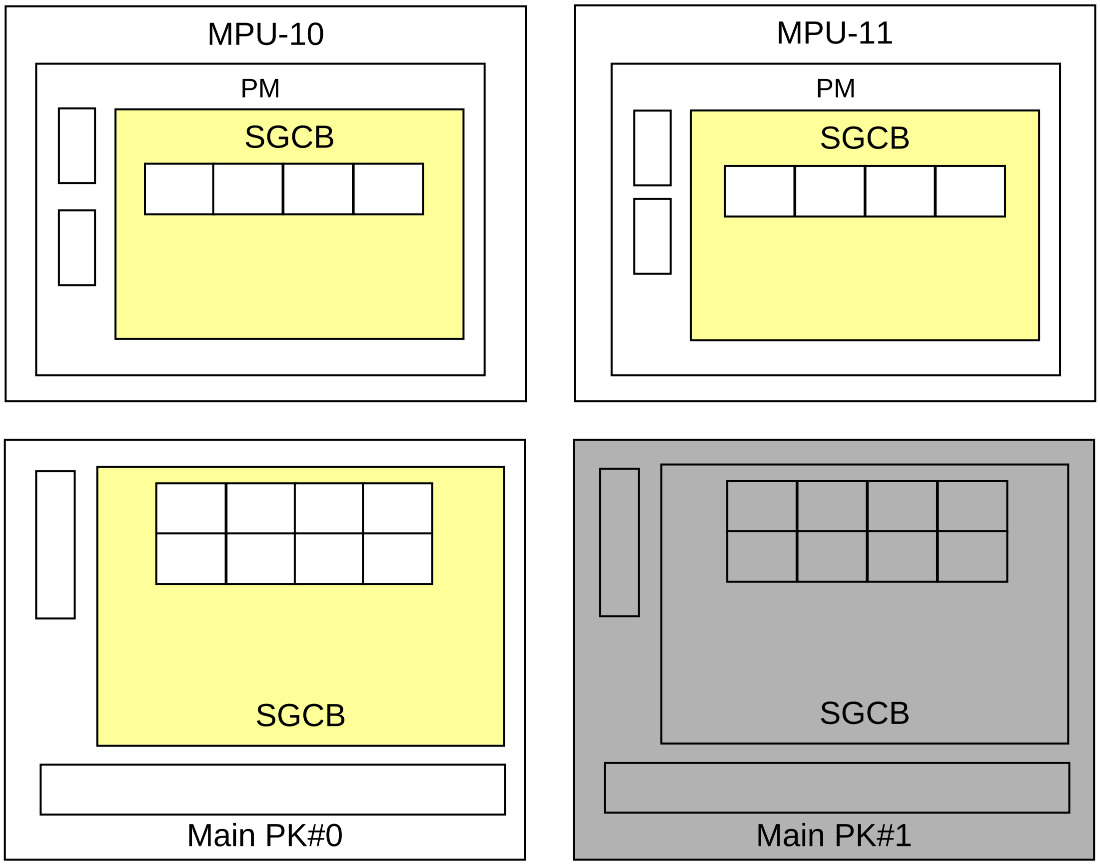


Fig. 3.9.4.4-1 Main Blade Replace (1)

Copyright © 2012, Hitachi, Ltd.

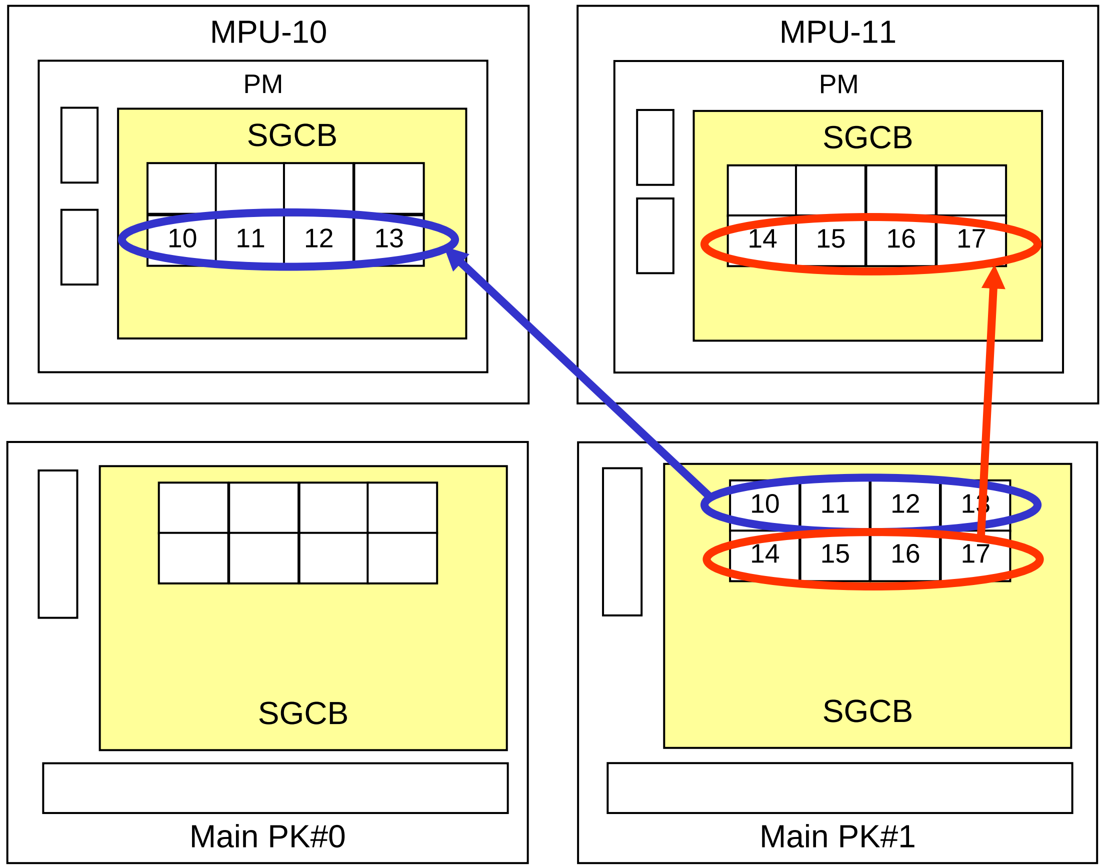


Fig. 3.9.4.4-2 Main Blade Replace (2)

Copyright © 2012, Hitachi, Ltd.

3.9.4.5 Ownership movement

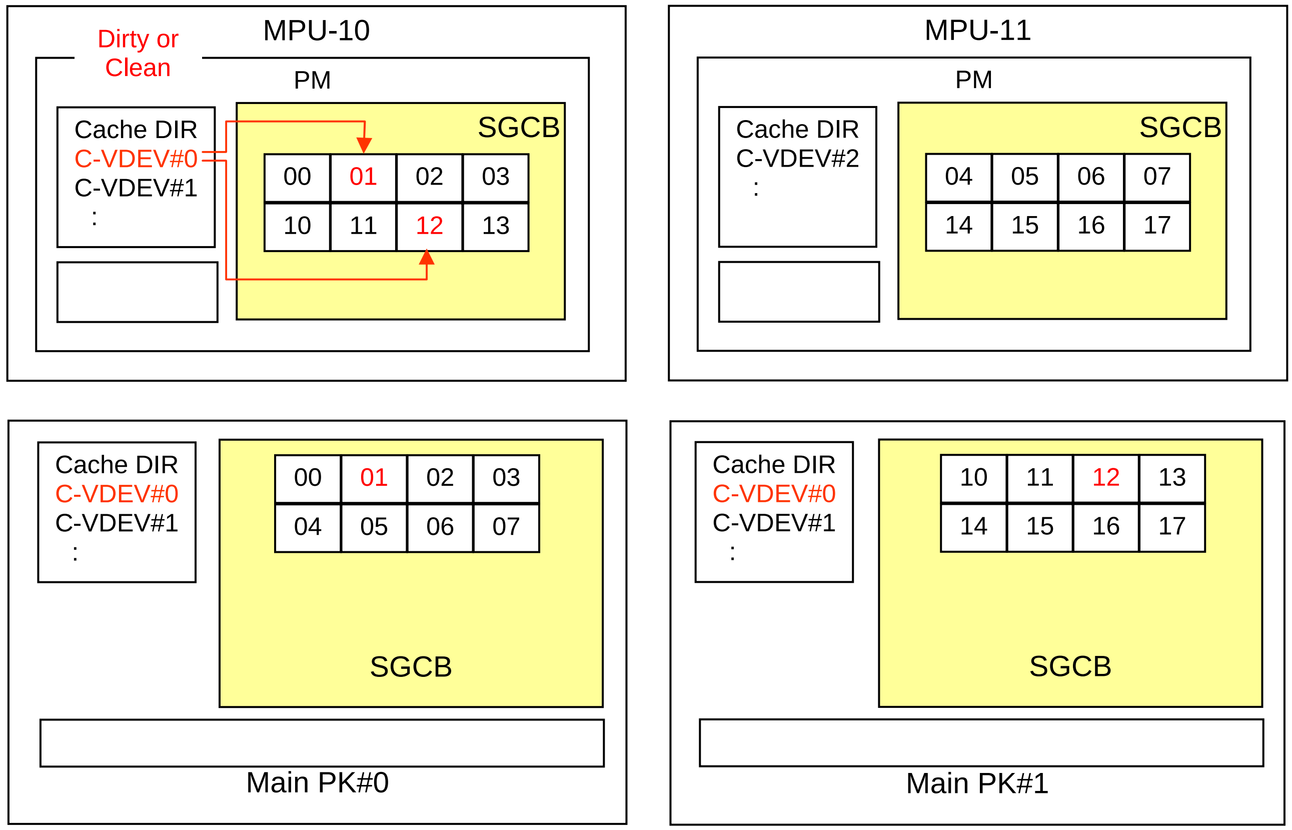


Fig. 3.9.4.5-1 Ownership movement (1)

Copyright © 2012, Hitachi, Ltd.

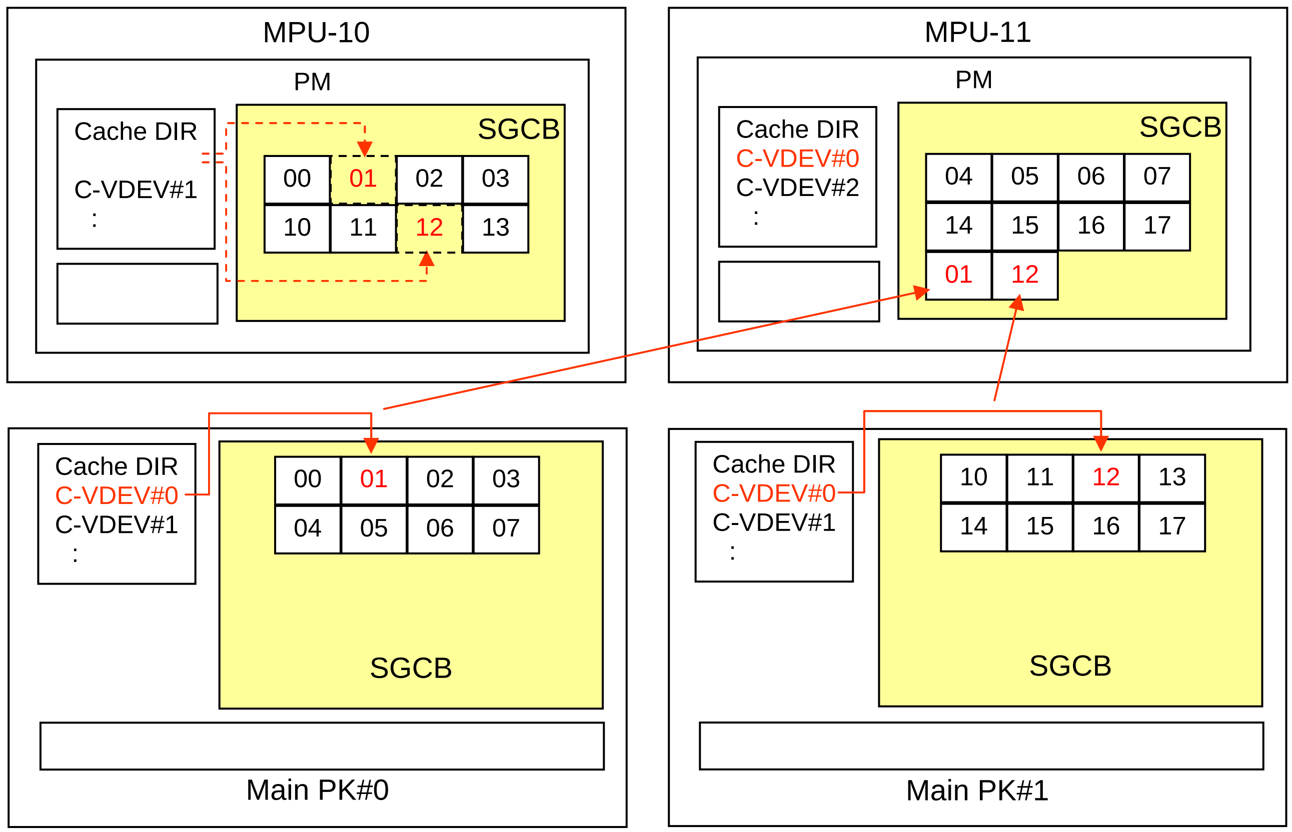


Fig. 3.9.4.5-2 Ownership movement (2)

Copyright © 2012, Hitachi, Ltd.

3.9.4.6 Cache Workload balance

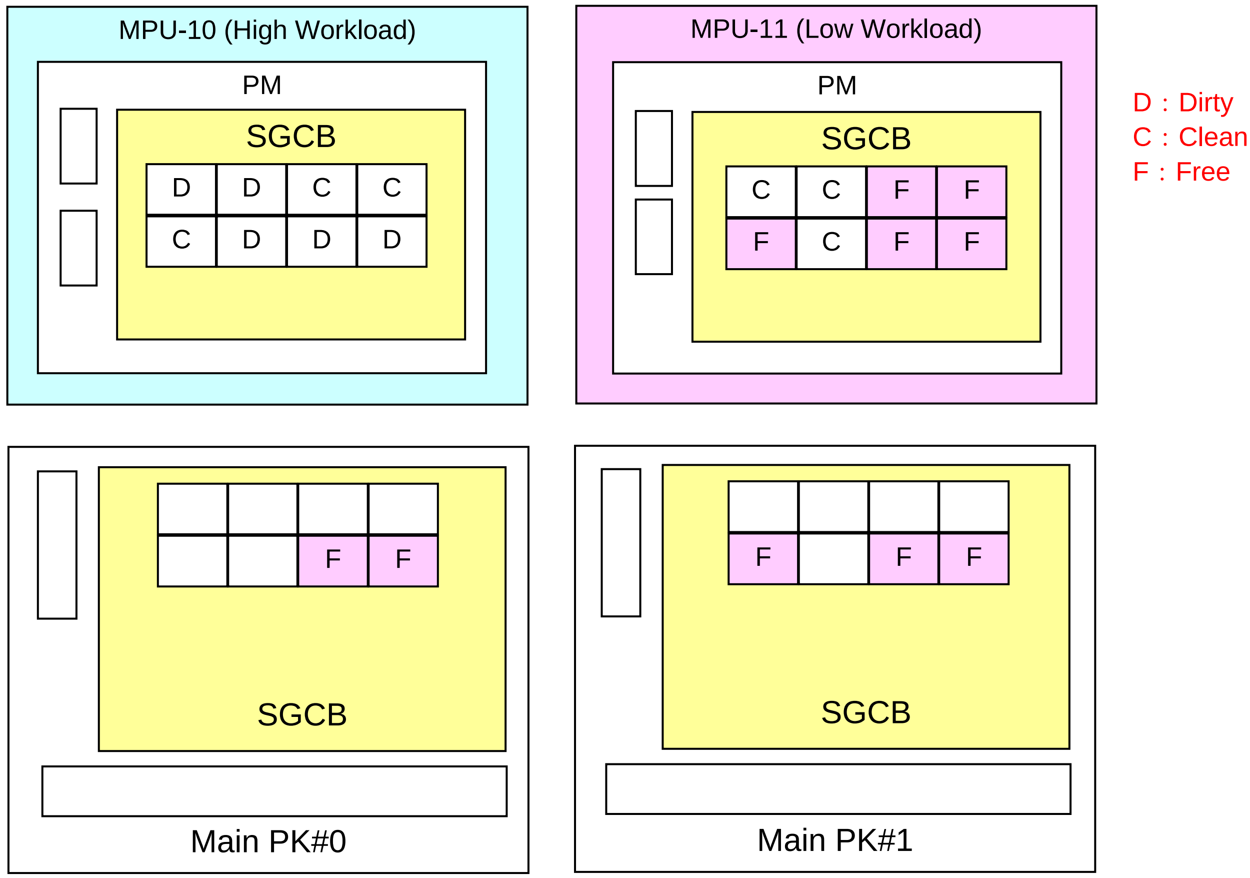


Fig. 3.9.4.6-1 Cache Workload balance (1)

Copyright © 2012, Hitachi, Ltd.

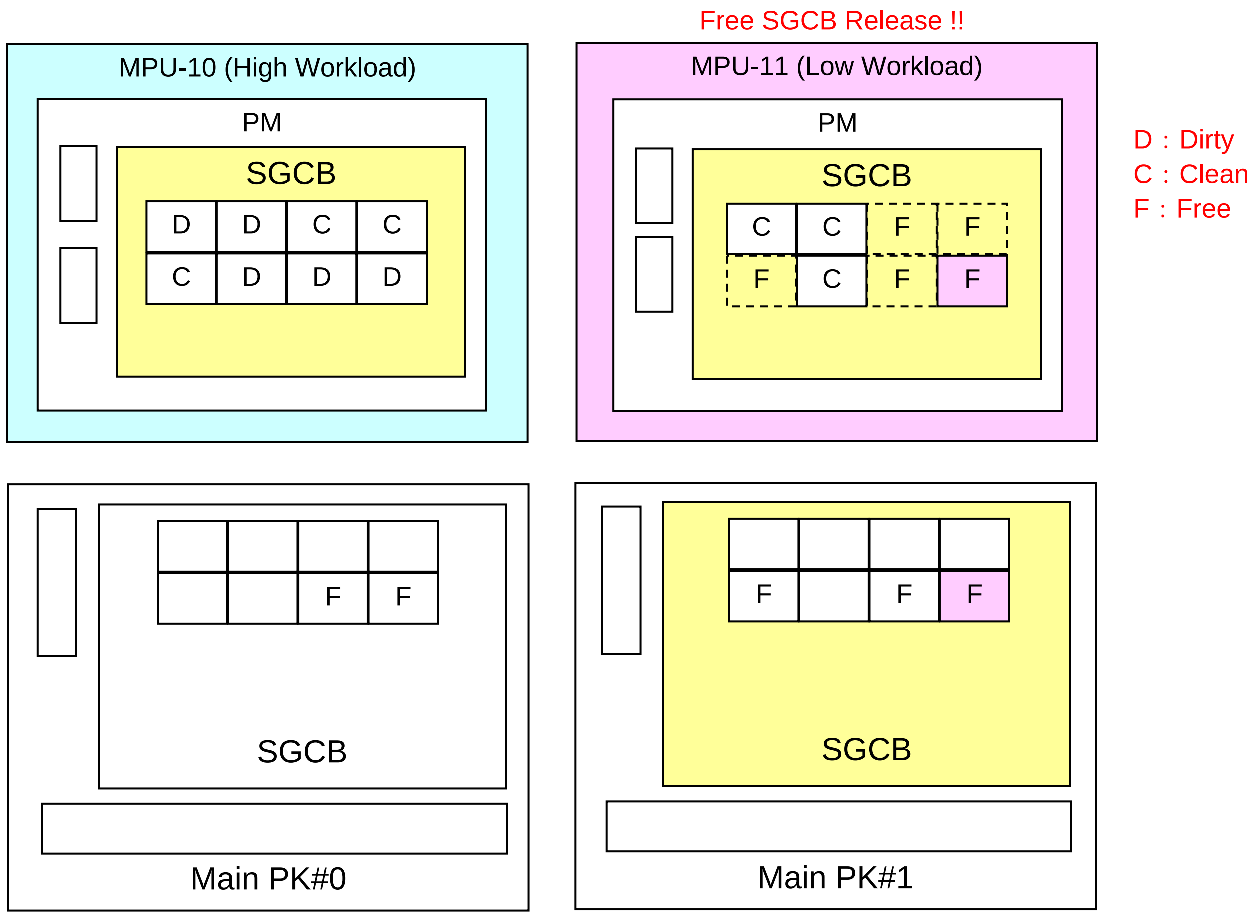


Fig. 3.9.4.6-2 Cache Workload balance (2)

Copyright © 2012, Hitachi, Ltd.

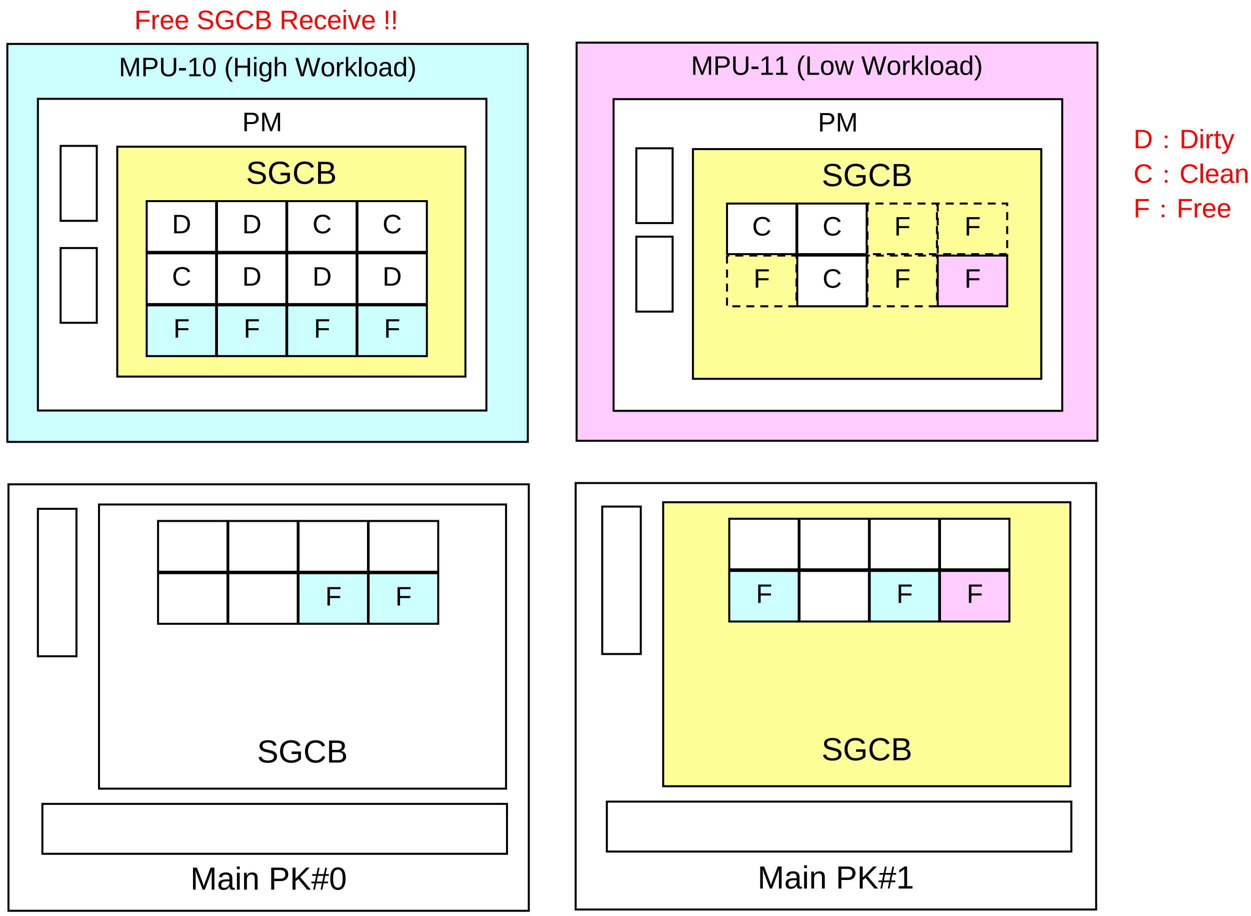


Fig. 3.9.4.6-3 Cache Workload balance (3)

Copyright © 2012, Hitachi, Ltd.

3.9.4.7 MP Blade Replace

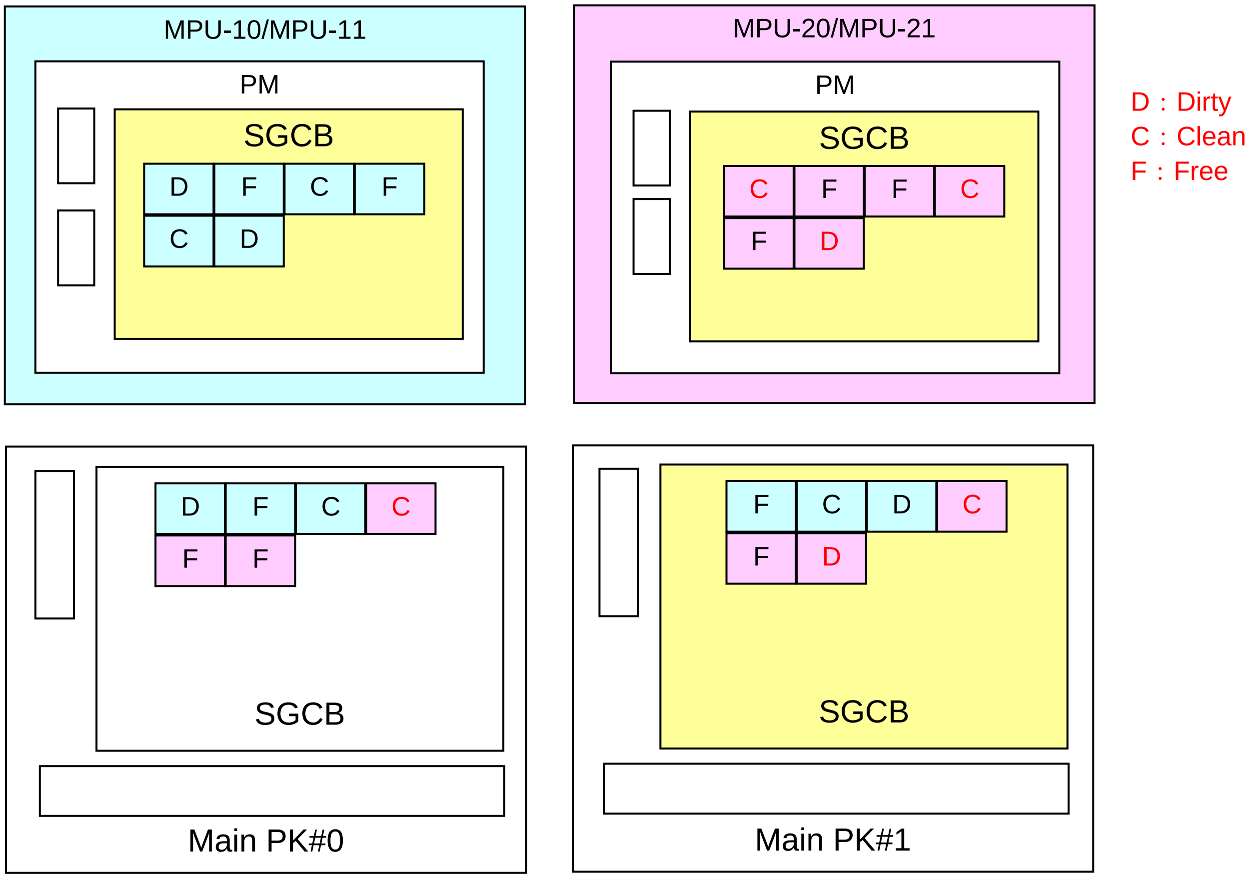


Fig. 3.9.4.7-1 MP Blade Replace (1)

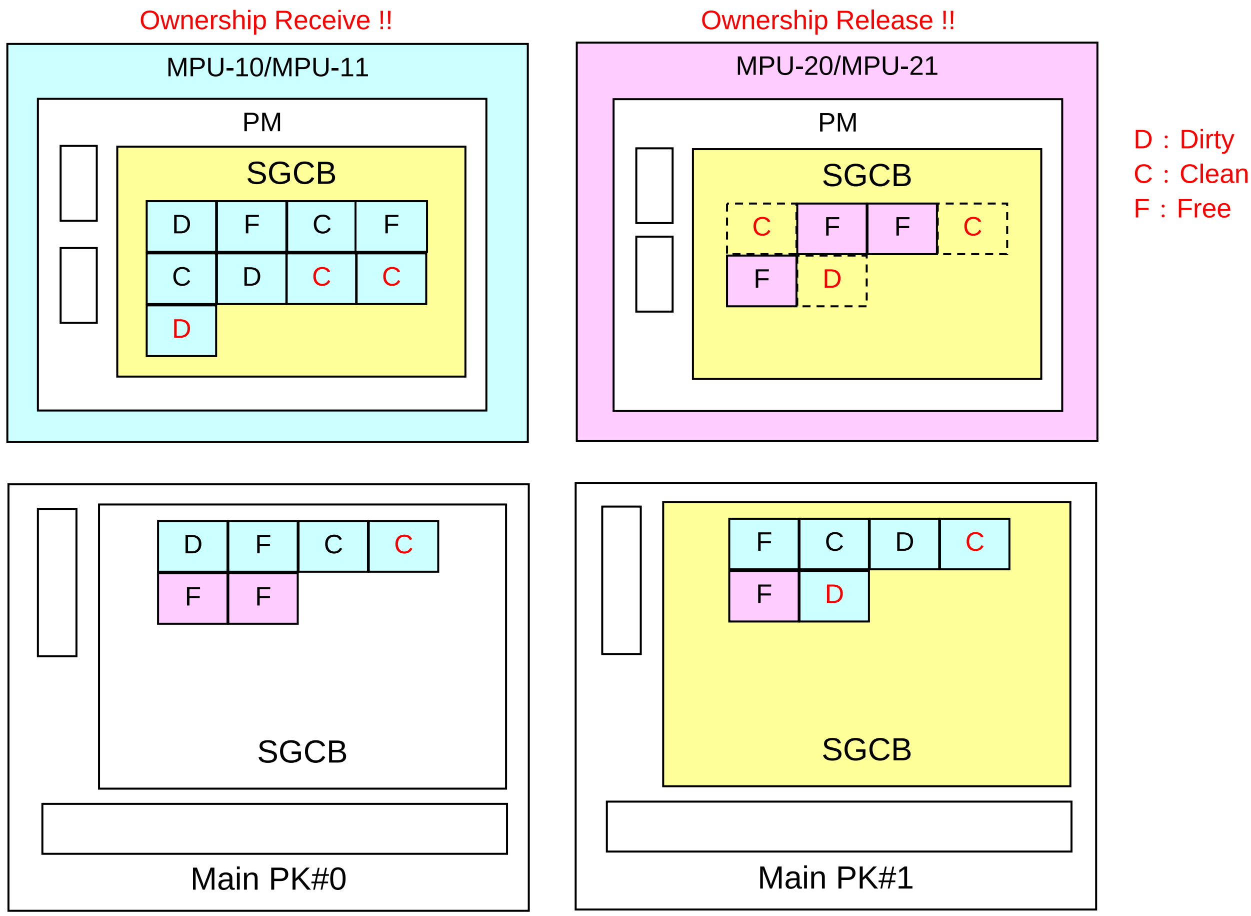


Fig. 3.9.4.7-2 MP Blade Replace (2)

 Copyright © 2012, Hitachi, Ltd.

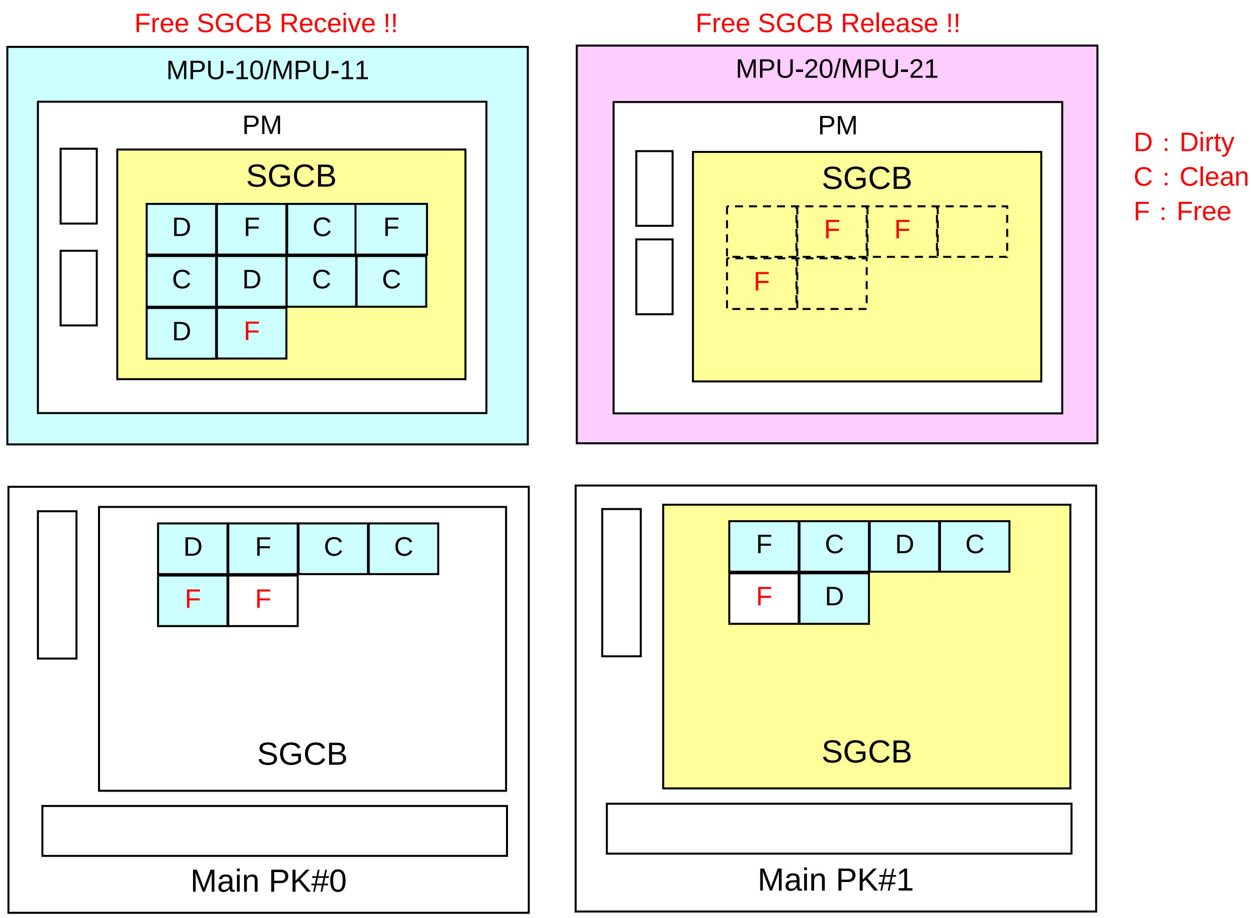


Fig. 3.9.4.7-3 MP Blade Replace (3)

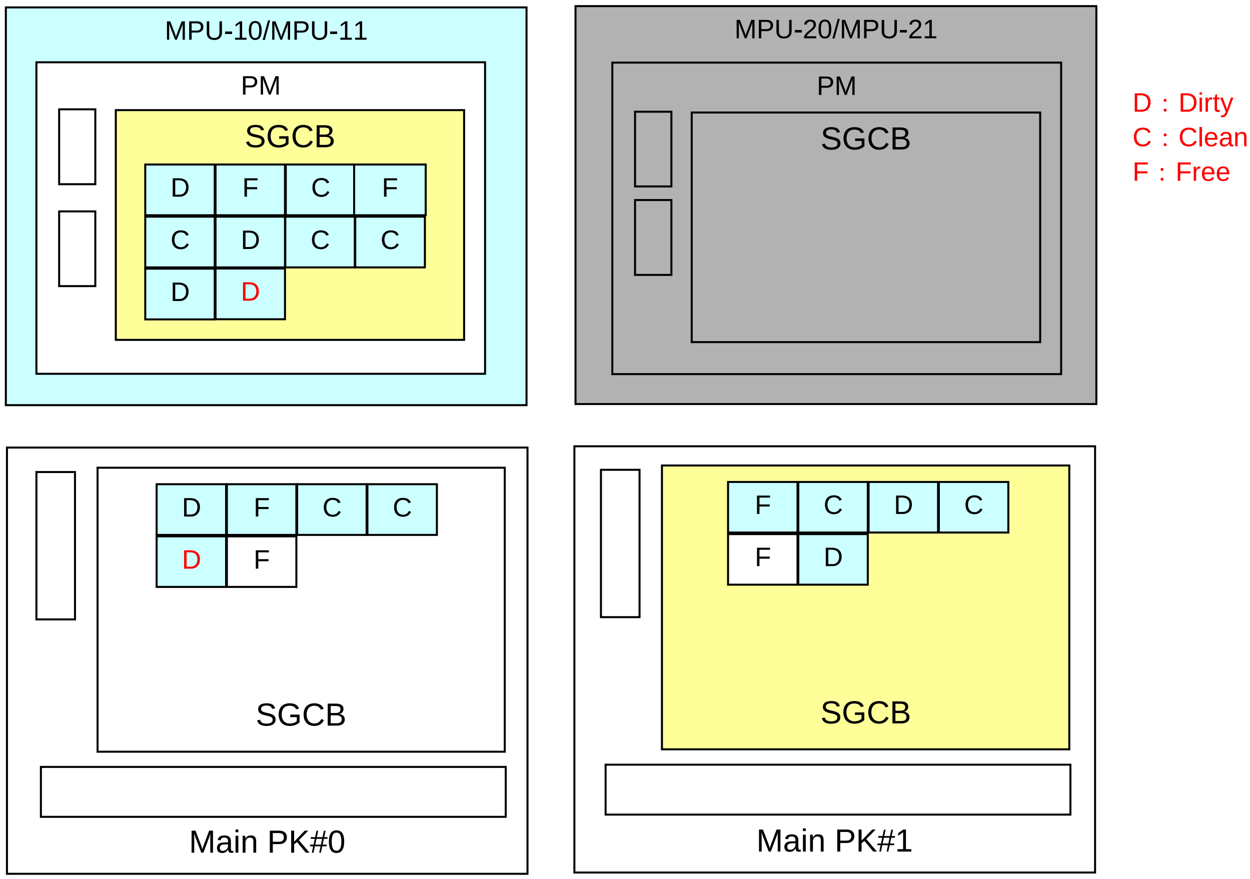


Fig. 3.9.4.7-4 MP Blade Replace (4)

 Copyright © 2012, Hitachi, Ltd.

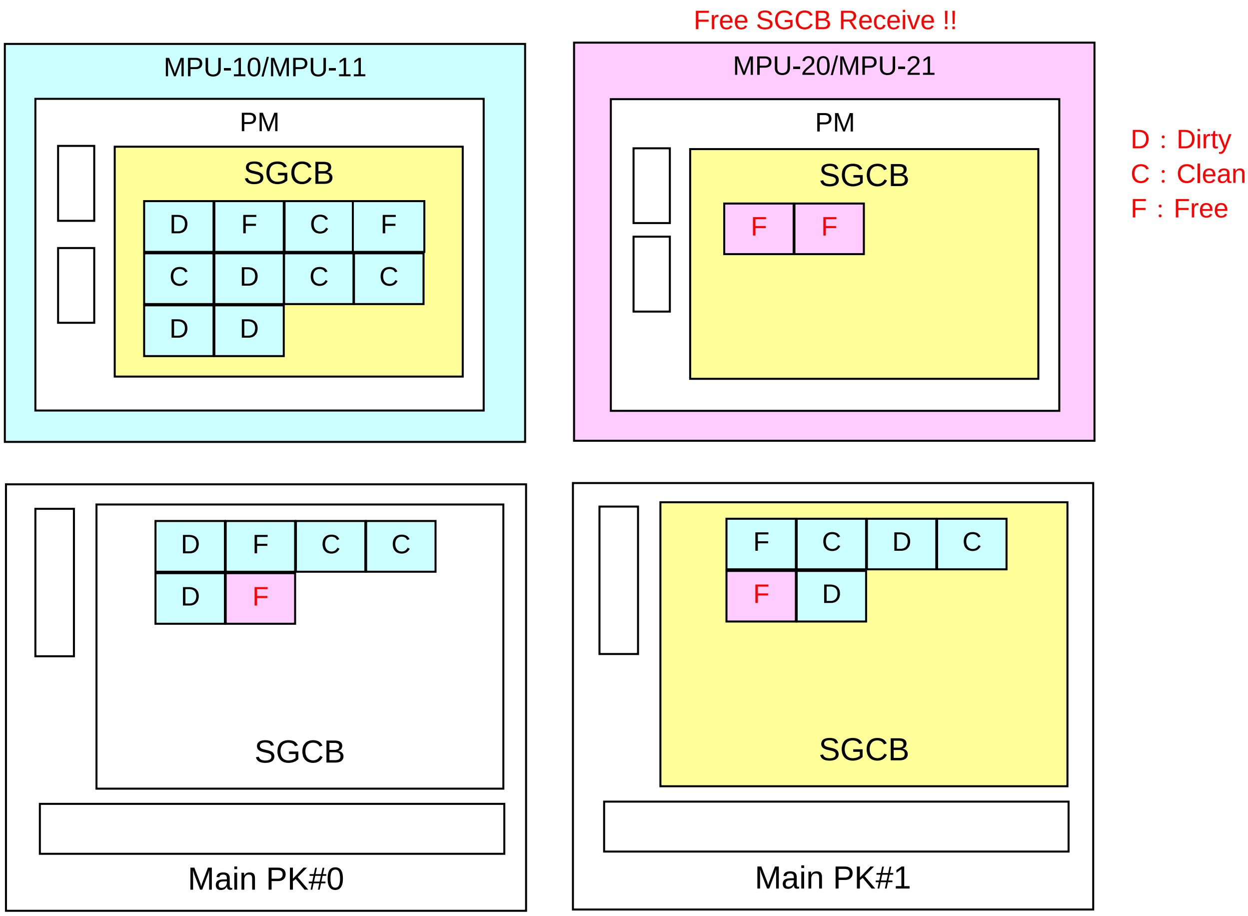


Fig. 3.9.4.7-5 MP Blade Replace (5)

 Copyright © 2012, Hitachi, Ltd.

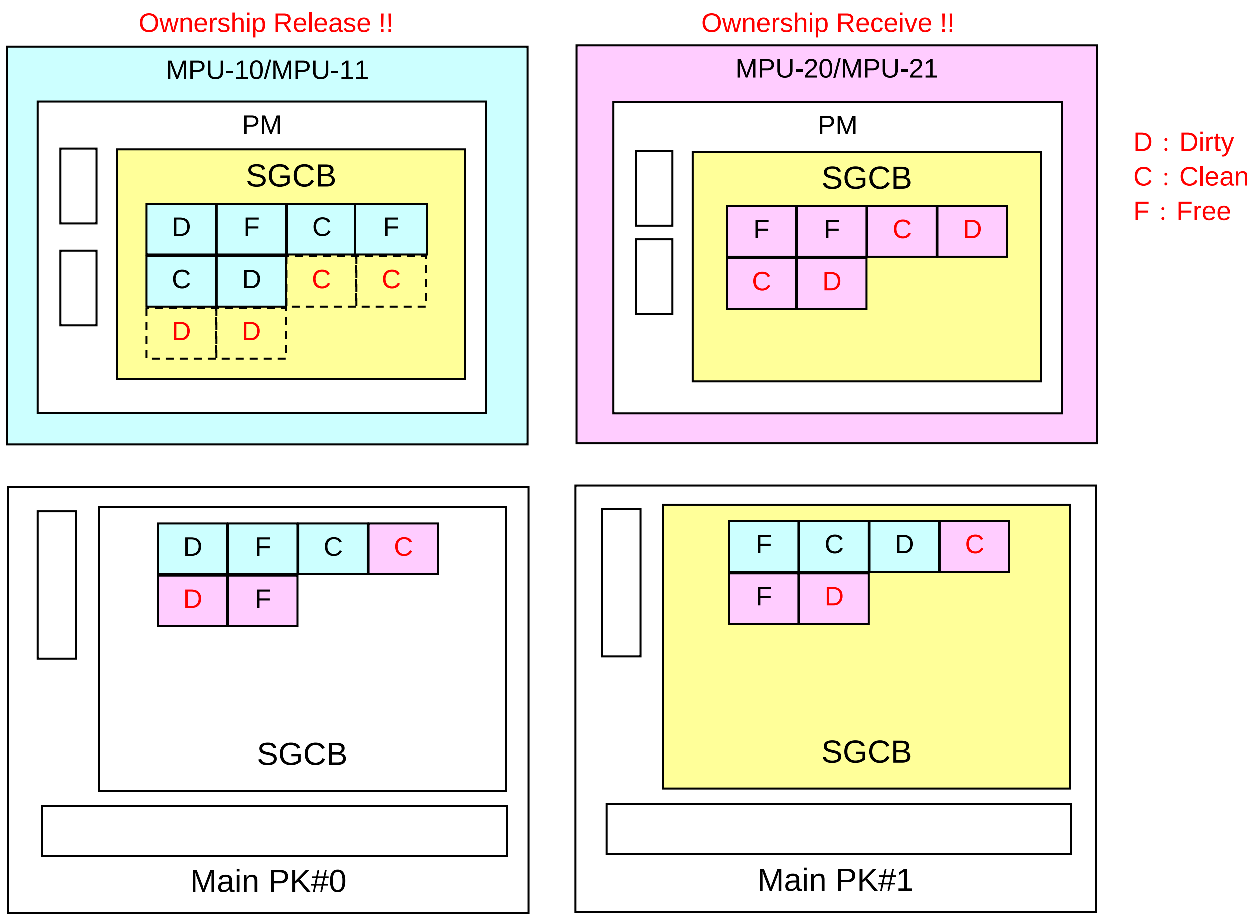


Fig. 3.9.4.7-6 MP Blade Replace (6)

 Copyright © 2012, Hitachi, Ltd.

3.9.4.8 Queue/Counter Control

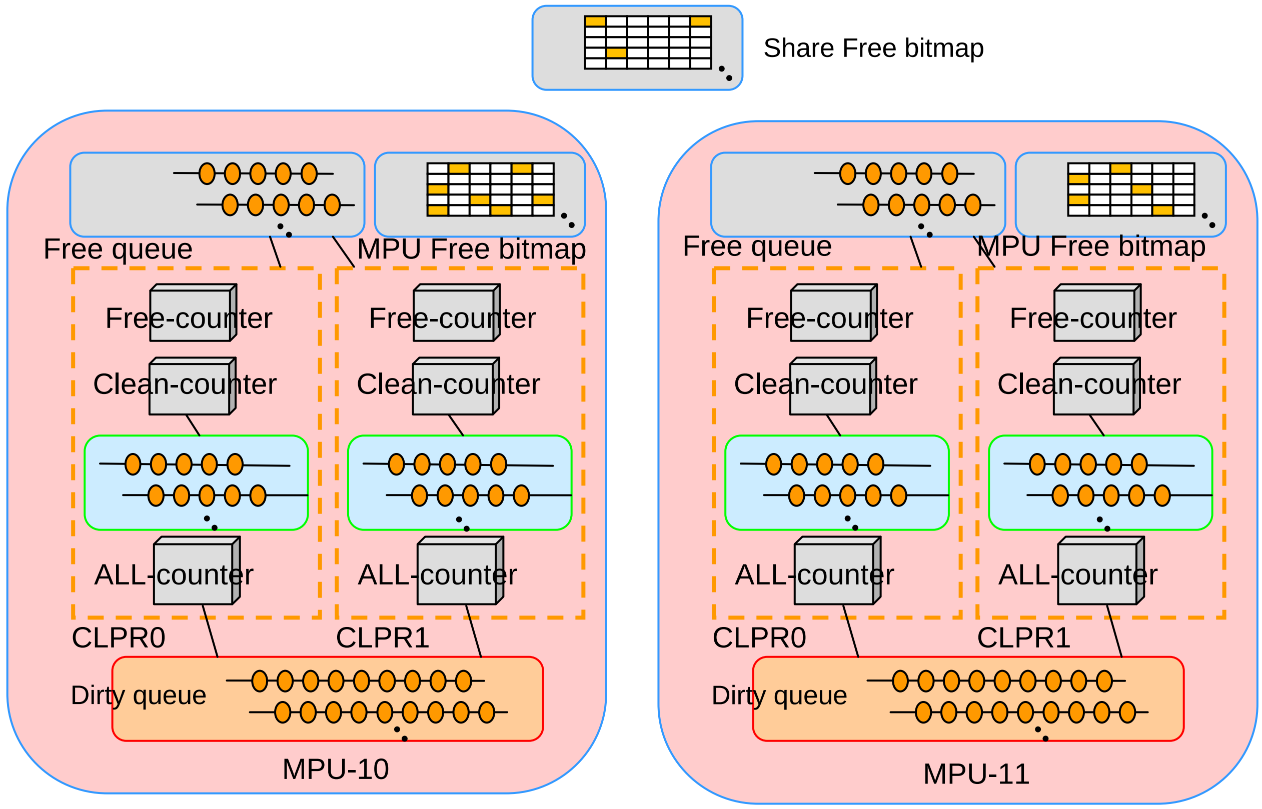


Fig. 3.9.4.8-1 Queue/Counter Control (1)

 Copyright © 2012, Hitachi, Ltd.

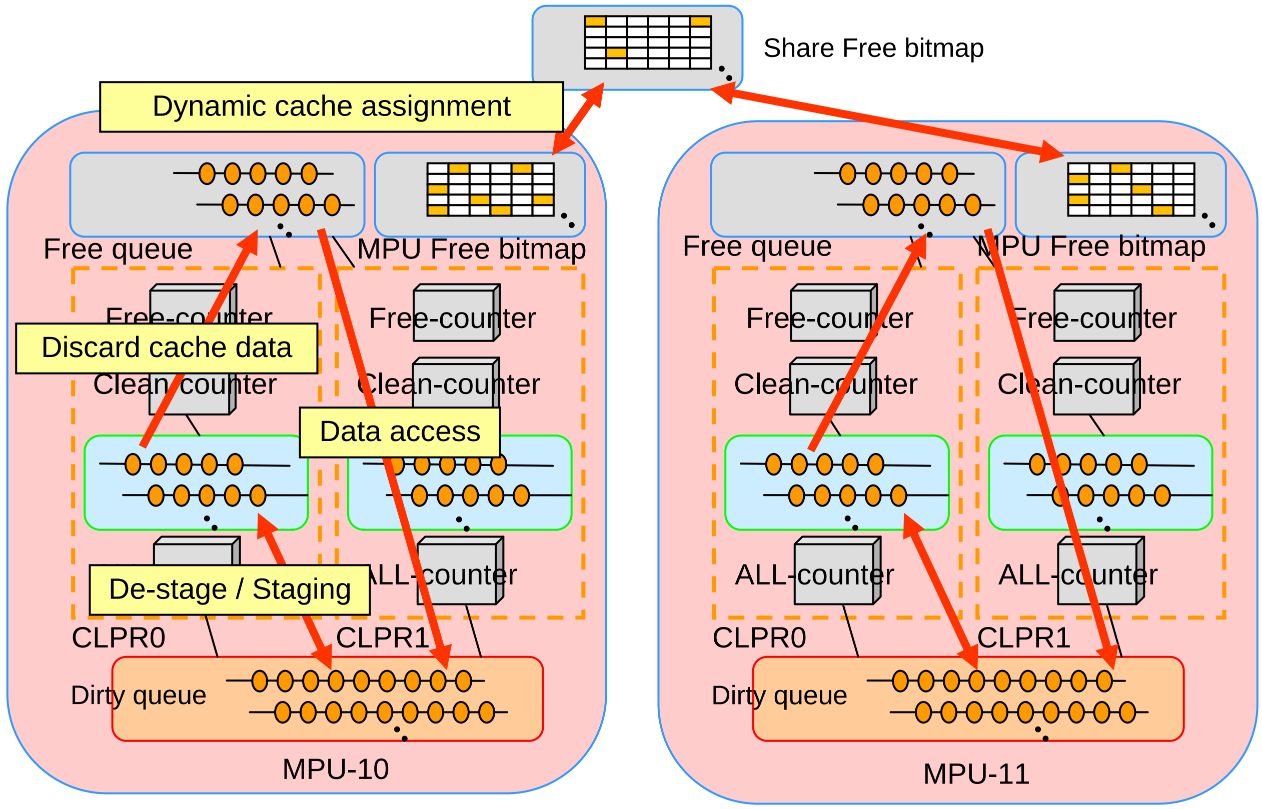


Fig. 3.9.4.8-2 Queue/Counter Control (2)

Copyright © 2012, Hitachi, Ltd.

3.10 TrueCopy

3.10.1 TrueCopy Components

(1) TrueCopy Components

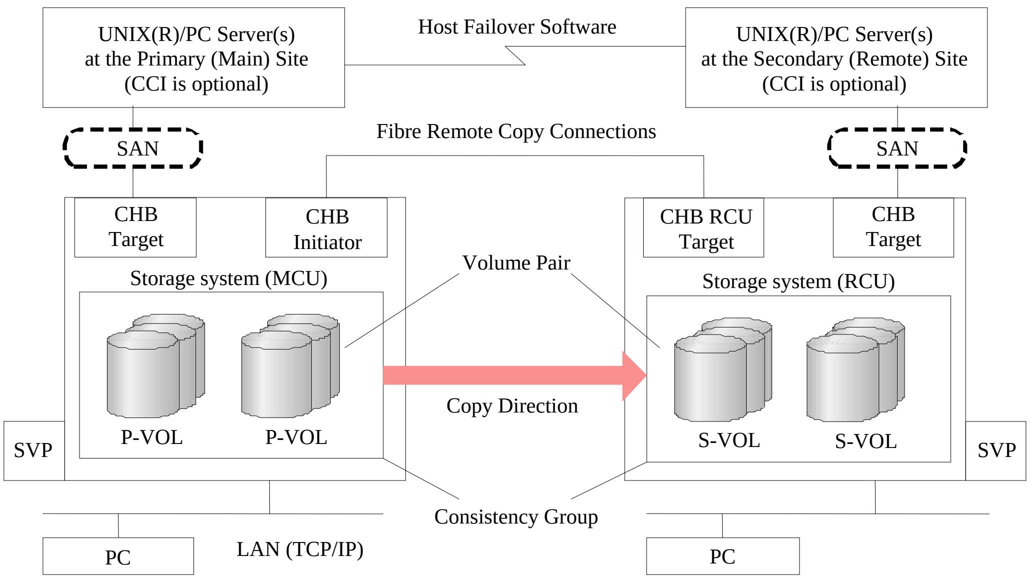


Fig. 3.10.1-1 TrueCopy Components for Fibre Channel Interface Connection

Copyright © 2012, Hitachi, Ltd.

(a) TrueCopy Volume Pair

An TrueCopy volume pair consists of two logical volumes, an P-VOL and an S-VOL, in different storage system.

An P-VOL is a primary volume. It can be read or written by I/O operations from host servers.

An S-VOL is a secondary or a mirrored volume.
Under control of the storage systems, contents of an P-VOL and updates from host servers are copied to an S-VOL. Read or write I/O operations from host server to S-VOLs are rejected.

NOTE: S-VOL Read/Write Mode
Read I/O operations to S-VOL of splitted pairs of TrueCopy is accepted. And if “S-VOL Write: Enable” was specified when pairsplit operation, read and write I/O operation is accepted.

The P-VOLs of the TrueCopy volume pairs and the S-VOLs of other TrueCopy volume pairs can be intermixed in one storage system.

(b) MCU and RCU

An **MCU** (main disk control unit) and an **RCU** (remote disk control unit) are disk control units in the storage systems to which the P-VOLs and the S-VOLs are connected respectively.

An MCU controls I/O operations from host servers to the P-VOLs and copy activities between the P-VOLs and the S-VOLs. An MCU also provides functions to manage TrueCopy status and configuration.

An RCU executes write operations directed by the MCU. The manner to execute write operations is almost same as that of I/O operations from host servers. An RCU also provides a part of functions to manage TrueCopy status and configuration.

Note that an MCU/RCU is defined on each TrueCopy volume pair basis. One disk control unit can operate as an MCU to control the P-VOLs and an RCU to control the S-VOLs.

(c) Remote Copy Connections

There Fibre channel interface of connection form.
At least two independent remote copy connections should be established between an MCU and an RCU.

(d) SVP and Web Console

An **SVP** provides functions to set up , modify and display TrueCopy onfiguration and status.

A **Web Console** is a personal computer compatible with the PC/AT. It should be connected to storage systems with an Ethernet network (TCP/IP). Several storage systems can be connected with one Ethernet network.

For Web Console, Hitachi provides only two software components, an TrueCopy application program and dynamic link library. Both of them require Microsoft Windows operating system. A personal computer, Ethernet materials and other software products are not provided by Hitachi.

Copyright © 2012, Hitachi, Ltd.

(e) Initiator Port

An **Initiator Port** (remote control port) is a Fibre Channel interface port to which an RCU is connected. Any Fibre Channel interface port of the storage systems can be configured as an Initiator Port.
But, as for the channel port of the host computer, it can't communicate. A path from the host computer must be connected with other Fibre Channel interface ports.

(f) RCU Target Port

An **RCU Target Port** (remote control port) is a Fibre Channel interface port to which an MCU is connected. Any Fibre Channel interface port of the storage systems can be configured as an RCU Target Port.
It can be connected with the channel of the host computer by the Fibre Channel switch.

Copyright © 2012, Hitachi, Ltd.

3.10.2 TrueCopy Hardware Requirements

(1) TrueCopy Supported models

Refer to Specifications of “THEORY OF OPERATION SECTION” for the Support models.

- CVS/DCR is able to define on the P-VOL and S-VOL.
- When T-VOL of Shadow Image is created as P-VOL of TrueCopy, the T-VOL must be “Split” state. If not so, a error will occur in creating TrueCopy pair.

(2) Web Station PC Requirements

An TrueCopy application software and dynamic link library require Microsoft Windows.

Copyright © 2012, Hitachi, Ltd.

(3) Distance between MCU and RCU

(a) Fibre channel interface connection
You must connect MCU and RCU with Optical Fibre cable.
With ShortWave (Optical Multi Mode), the longest cable is 500 m. The longest cable is 10 km with LongWave (Optical Single Mode).
By connecting Switch, the longest cable for ShortWave is 1.5 km, and 30 km for LongWave.
But the Switch can be connected with a maximum of two steps.

Channel Extender Connects MCU and RCU with no distance restriction.

In case of a direct connection between MCU and RCU, each Fibre channel port topology must be “Fabric:Off and FC-AL”.

In case of via FC-Switch connection between MCU and RCU, each Fibre channel port topology must be set the same as for the closest FC-Switch’s topology.
(Eg.) “Fabric:On and FC-AL” or “Fabric:On and Point-to-Point” or “Fabric:Off and Point-to-Point”

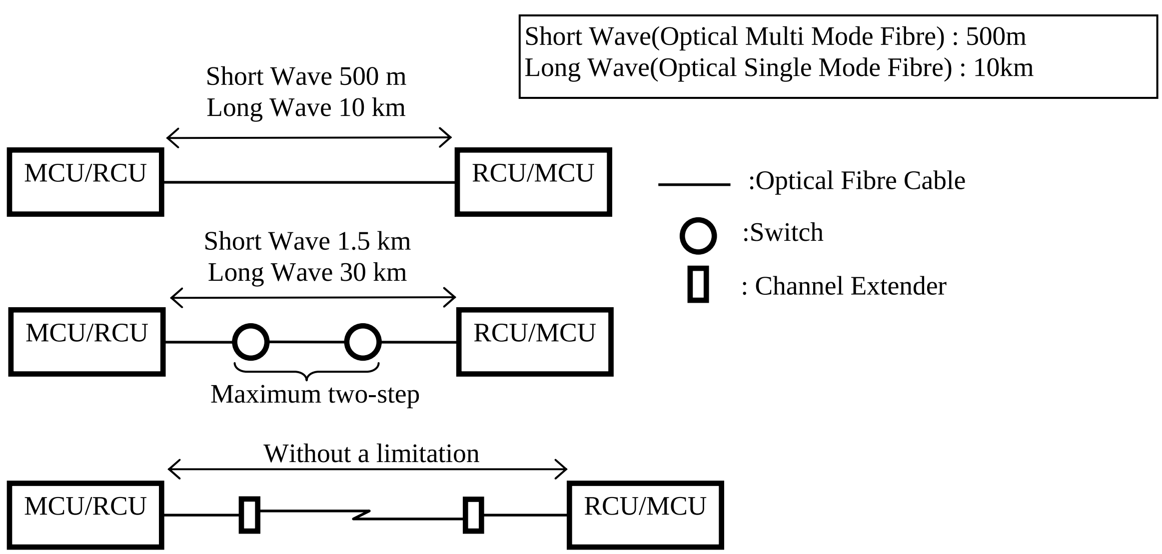


Fig. 3.10.2-1 Distance between Storage systems of Fibre channel interface connection

Copyright © 2012, 2013, Hitachi, Ltd.

(4) Recommendation of MIH time and TrueCopy configuration

Recommendation of MIH time is 60 sec. for TrueCopy. In addition that, MIH time needs to be set with consideration of the following factors.
• The number of paiS-VOLumes
• Cable length between MCU and RCU
• Volume status (Initial copy status)
• Maintenance operation pending

(5) TrueCopy available CU image
CU#0 ~ CU#254 (In hexadecimal, CU#FE)

(6) Restriction for Connecting with Former Model of Storage System
HUS VM can connect HUS VM, VSP, or USP V/VM. When you connect HUS VM with VSP or USP V/VM, contact the technical support division for the available microcode version of each model.

Copyright © 2012, Hitachi, Ltd.

AppendixA: TrueCopy Installation check list

Table 3.10.2-1 TrueCopy Installation Check List

No.	Item	Check
1	MCU/RCU emulation type must be correct.	
2	P-VOL/S-VOL emulation type must be correct.	
3	Initiator port must be set.	
4	Fiber channel cable between MCU and RCU must be connected.	
5	Fiber channel cable test between MCU and RCU must be executed.	

Copyright © 2012, Hitachi, Ltd.

3.10.3 TrueCopy Theory of Operations

(1) TrueCopy Copy Activities
TrueCopy executes two kinds of copy activities, initial copy and update copy.

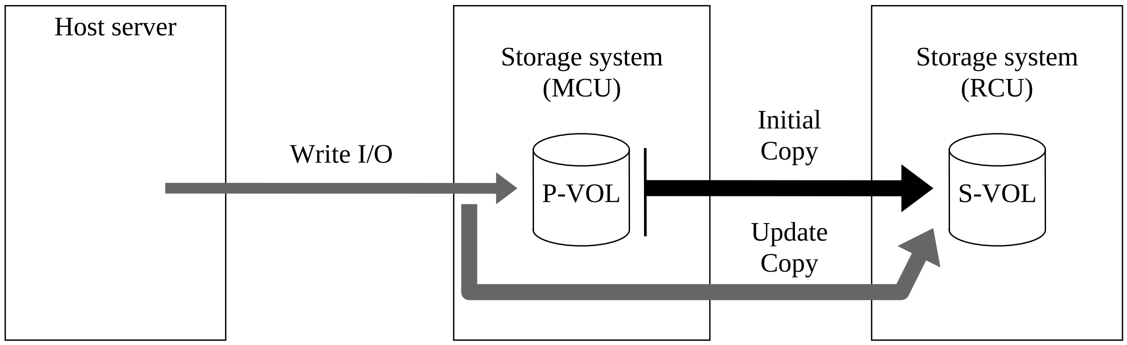


Fig. 3.10.3-1 TrueCopy Copy Activities

(a) Initial Copy

Responding to an Create TC volume Pair operation from SVP/Web Console or RAID Manager command, TC begins initial copy. And all data is copied from P-VOL to S-VOL.

“No copy” can be specified as a parameter to the initial copy. When “no copy” is specified, TC will complete an Create TC Volume Pair operation without copying any data. An operator or a system administrator should be responsible for ensuring that data on the P-VOL and the S-VOL is already identical.

(b) Controlling Initial Copy

Number of tracks copied by one initial copy activity can be specified by an SVP/Web Console or a RAID Manager command.

Number of volume pairs for which the initial copy are concurrently executed and priority of each volume pair can be specified from an SVP/Web Console.

(c) Update Copy

Responding to the write I/O operations from the host servers, TC copies the data updated by the write I/O operation to the S-VOL.

The update copy is a synchronous remote copy. An MCU starts the update copy after responding only channel-end status to the host servers channel, and sends device-end status after completing the update copy.

If many consecutive records are updated by single CCW chain which does not use locate record command, the third condition above may cause the significant impact on performance.

(d) Special Write Command for Initial Copy and Update Copy

In order to reduce overhead by the copy activities, TC uses a special write command which is allowed only for copy activities between the storage systems. The single write command transfers control parameters and an FBA formatted data which includes cunsecutive updated data in track.

(2) TrueCopy Read I/O Operations

Responding to read I/O operations, an MCU transfers the requested records form an P-VOL to a host server. Even if reading records from the P-VOL is failed, the S-VOL is not automatically read for recovery. The redundancy of the P-VOL itself provided by RAID1, RAID5 or RAID6 technique would recover the failure.

(3) TrueCopy Volume Pair Status

All volumes in a storage system are in one of the states shown in Table 3.10.3-1.

Status of the P-VOLs or the S-VOLs are kept by the MCU and the RCU respectively. The MCU is responsible to keep status of the S-VOLs identical to status of the P-VOLs. However, in the case of communication failure between the MCU and the RCU, they could be different.

From an Web Console or by using an appropriate command for RAID Manager,

Table 3.10.3-1 TrueCopy Volume Status

Status	Description
SMPL	This volume does not belong to TC volume pair. When the initial copy is started by an Add Pair operation, the volume is changed to “COPY” state.
COPY	The initial copy is in progress. Data on TC volume pair is not fully identical. When completing the initial copy, the volume will be changed to “PAIR” state.
PAIR	Volumes in TC volume pair are synchronized. All updates from the host servers to the P-VOL are duplicated to the S-VOL.
PSUS/PSUE	Volumes in TC volume pair are not synchronized. • When the MCU can not keep synchronization between TC volume pair due to, for example, failure on the update copy, the MCU will put the P-VOL and the S-VOL in this state. • When the MCU or the RCU accepts a Split operation from an SVP/Web Console, the P-VOL and the S-VOL will be put in this state. • When the RCU accepts the Delete Pair operation from the SVP/Web Console, the MCU will detect the operation and put the P-VOL in this state.

 Copyright © 2012, Hitachi, Ltd.

Table 3.10.3-2 TrueCopy Volume Status - Sub-status of Suspended Volume

Cause of Suspension	Description
P-VOL by Operator	The Suspend operation with "P-VOL failure" option was issued to the P-VOL. This cause of suspension is defined only for the P-VOLs.
S-VOL by Operator	The Suspend operation with "S-VOL" option was issued to the P-VOL or the S-VOL. This cause of suspension is defined for both the P-VOLs and the S-VOLs.
by MCU	The RCU received a request to suspend the S-VOL from an MCU. This cause of suspension is defined for only the S-VOLs.
by RCU	The MCU detected an error condition of the RCU which caused TC volume pair to be suspended. This cause of suspension is defined only for the P-VOLs.
Delete Pair to RCU	The MCU detected that S-VOL had been changed to "SMPL" state by the Delete Pair operation. This cause of suspension is defined only for the P-VOLs.
S-VOL Failure	The MCU detected an error condition on the communication between the RCU or I/O error on the update copy. This cause of suspension is defined only for the P-VOLs. The cause of suspension of the S-VOLs are usually set to "by MCU" in this situation.
MCU IMPL	The MCU could not find valid control information in its non-volatile memory during its IMPL procedure. This situation may occur after the power supply failure.
Initial Copy Failed	The volume pair was suspended before completing the initial copy. Even if no write I/O has been issued after being suspended, the data in the S-VOL is not completely identical to the P-VOL.

 Copyright © 2012, 2013, Hitachi, Ltd.

3.10.4 TrueCopy Control Operations

This section describes TrueCopy control operations from a Web Console.

(1) Add RCU Operation

(a) Fibre channel interface connection

The following parameters are necessary to register RCU as a Fibre Channel connection.

Port Type	Fibre: Fiber channel interface is used for the connection of MCU and RCU.
Controller ID	Set it up with ‘19’ when RCU is HUS VM, set it up with ‘6’ when RCU is VSP, or set it up with ‘5’ when RCU is USP V/VM.
RIO MIH Time	A data transfer complete waiting time to RCU from MCU. Usual: 15[Sec]. Avail. range: 10[Sec] ~ 100[Sec]
MCU Port	An Initiator port of the storage system which set up a logic pass. You must set up a Fibre Channel interface port in Initiator port before this operation.
RCU Port	The Fibre Channel interface port of the place of the connection. You must specify a RCU target port.

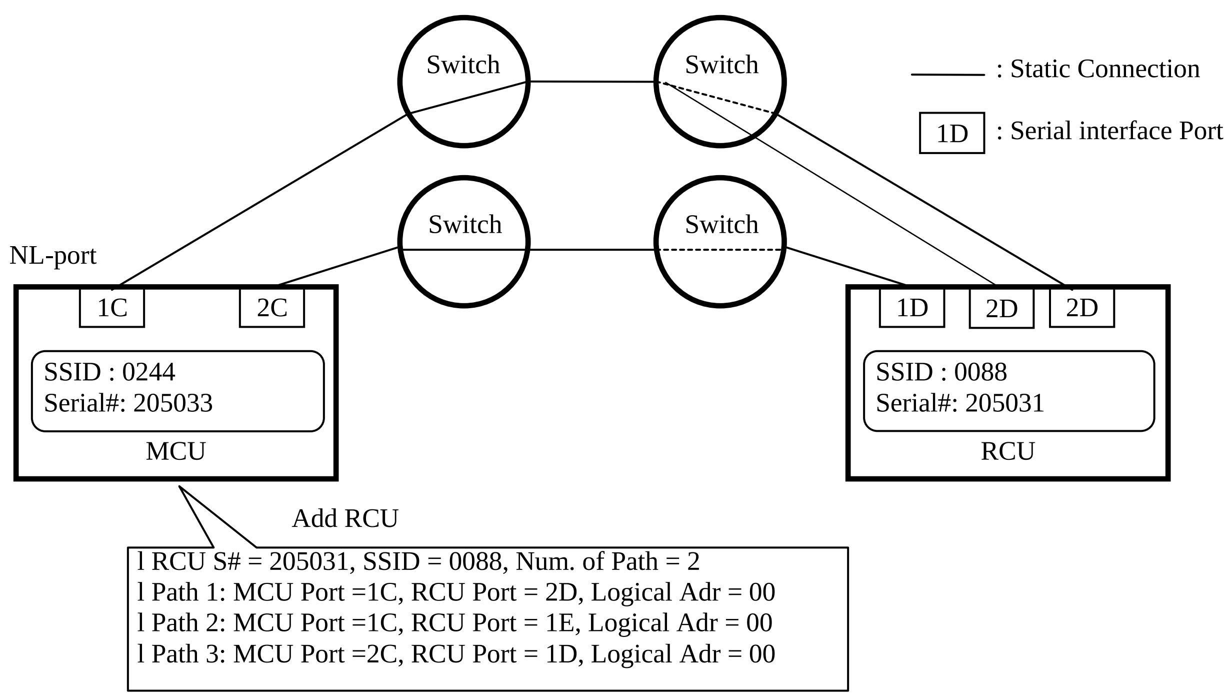


Fig. 3.10.4-1 Add RCU Operation

 Copyright © 2012, Hitachi, Ltd.

The following parameters modify the Remote Copy options which will be applied to all Remote Copy volume pairs in this storage system.

Minimum Paths	When the MCU blocks the logical path due to communication failure, if the number of remaining paths becomes less than the number specified by this parameter, the MCU will suspend all of the Remote Copy volume pairs. The default value is set to “1”. If the installation requirements prefers the storage system I/O performance to the continuation of Remote Copy, value between “2” and the number of the established logical paths can be specified.
Maximum Initial Copy Activities	It specifies how many TC initial copies can be simultaneously executed by the MCU. If more Remote Copy volume pairs are specified by an Add Pair operation, the MCU will execute the initial copy for as many volumes as specified by this parameter. The initial copy for otheS-VOLumes is delayed until one of the initial copies is completed. This parameter can control the performance impact caused by the initial copy activity. NOTE: Default value of this parameter is “64”.

Note that these parameters will be applied to ALL RCUs registered to the MCU. If different parameters are specified, the last parameter will be applied.

Copyright © 2012, Hitachi, Ltd.

(2) Edit Path Operation

An Edit Path operation makes the MCU add/delete the logical path to the registered RCU.

To add a logical path, the same path parameters as an Add RCU operation are required. The added logical path will be automatically used to execute the copy activities.

When deleting a logical path, pay attention to the number of remaining logical paths. If it becomes less than the number specified by “Minimum Paths”, Remote Copy volume pair could be suspended.

(3) RCU Option Operation

An RCU Option operation modifies the Remote Copy options described in “3.10.4(1) Add RCU operation”.

(4) Delete RCU Operation

A Delete RCU operation makes the MCU delete the specified RCU from RCU registration. All logical paths to the specified RCU will be removed.

If some volumes connected to the specified RCU are active S-VOLs, this operation will be rejected. All S-VOLs must be deleted by a Delete Pair operation before a Delete RCU operation.

(5) RCU Status Operation

An RCU Status operation makes the MCU display the status of RCU registration. It also provides the current status, time of registration and time of changing status for each logical path.

The current status of each logical path is defined as follows:

Normal	This logical path has been successfully established and can be used for the Remote Copy activities.
Initialization Failed	The link initialization procedure between the RCU is failed. It occurred due to Missing physical path connection between MCU and RCU, or connecting MCU with HOST as RCU.
Resource Shortage (RCU)	Establish Logical Path link control function has been rejected by the RCU. All logical path resources in the RCU might be used for other connections.
Serial Number Mismatch	The serial number of the control unit which is connected to this logical path does not match to the serial number specified by “RCU S#” parameter.

Copyright © 2012, Hitachi, Ltd.

(6) Paircreate Operation

A Paircreate Pair operation makes the MCU establish a new Remote Copy volume pair. It also provides function to modify the Remote Copy options which will be applied to the selected Remote Copy volume pair.

To establish Remote Copy volume pair, following parameters are required:

RCU	The disk control unit which controls the S-VOL of this Remote Copy volume pair. It must be selected from RCUs which have already been registered by Add RCU operations.
S-VOL	Device number of the S-VOL.
Priority	Priority (scheduling order) of the initial copy for this volume pair. When the initial copy for one volume pair has been terminated, the MCU selects and start the initial copy for anotheS-VOLume pair which has the lowest value of this parameter. For the Paircreate operations, the value “1” through “256” can be specified. For establishing TC volume pair by TSO command or DSF command for PPRC, “0” is implicitly applied to. “0” is the highest priority, “256” is the lowest, and default value for the Paircreate operation is “32”. For the volume pairs to which the priority has been specified, the MCU prioritizes the volume pairs in the arrival order of the Paircreate operations or TSO/DSF commands. If the MCU are performing the initial copy for the number of volume pairs, as much as the value of “maximum initial copy activities”, and accepts further Add Pair operation, the MCU does not start other initial copy until one of the copy being performed will be completed. NOTE: When a time out occurs in this operation, a schedule may not be done as the priority parameter. The cause of the time-out is thought the problem of the configuration of DKC or Remote-copy connection path. Confirm configuration. After that, cancel a pair, and re-establish a pair.
Operation Mode	It specifies what kind of remote copy capability should be applied to this volume pair.
Initial Copy	It specifies what kind of initial copy activity should be executed for this TC volume pair. The kind of the initial copy can be selected out of: - “Entire Volume” specifies that all data should be copied. - “None” specifies that the initial copy does not need to be executed. The synchronization between volume pair must have been ensured by the operator.

Copyright © 2012, Hitachi, Ltd.

Remote Copy option parameters which will be applied to this Remote Copy volume pair are as follows:

Initial Copy Pace	It specifies how many tracks should be copied at once by the initial copy. From “1 Track” to “15 Tracks” can be specified. When number of track are many elapsed time to complete the initial copy becomes shorter, however, the storage system I/O performance during the initial copy could become worse. NOTE: Even if four or more tracks is specified, the number or tracks that is copied to S-VOL is restricted to four tracks.
P-VOL Fence Level	It specifies by what conditions the P-VOL will be fenced (the MCU will reject the write I/O operations to the P-VOL). - “S-VOL Data”: The P-VOL will be fenced when the MCU can not successfully execute the update copy. - “S-VOL Status”: The P-VOL will be fenced when the MCU can not put the S-VOL into “PSUS/PSUE” state. If status of the S-VOL is successfully changed to “PSUS/PSUE” the subsequent write I/O operations to the P-VOL will be permitted. - “Never”: The P-VOL will never be fenced. The subsequent write I/O operations after the TC volume pair has been suspended will be permitted.

Copyright © 2012, Hitachi, Ltd.

(7) Delete Pair Operation

A Delete Pair operation makes the specified Remote Copy volume pair being terminated. It can be operated on either the MCU or the RCU.

- When operated on the MCU, both the P-VOL and the S-VOL will be put into the “SMPL” state.
- When operated on the RCU, only the S-VOL will be put into the “SMPL” state. The P-VOL will be suspended when the MCU detects this operation. To complete deleting this volume pair, the MCU requires another Delete Pair operation.

When the MCU accepts this operation and it can not communicate with the RCU, this operation will be rejected. “Delete Pair by Force” option can make the MCU complete this operation, even if it can not communicate with the RCU.

(8) Split Pair Operation

A Split Pair operation makes the MCU or the RCU suspend the specified Remote Copy volume pair.

The option parameters for this operation are as follows:

P-VOL Failure	The subsequent write I/O operations to the P-VOL will be rejected regardless of the fence level parameter. This option can be selected only when operating on the MCU. This option is valid for only TC volume pairs.
S-VOL	For TC volume pairs. This option can be accepted by the MCU and the RCU.

(9) Pair Option Operation

A Pair Option operation modifies the Remote Copy option parameters which has been applied to the selected Remote Copy volume pair. Refer to “3.10.4(6) Add Pair Operation” for the option parameters.

Copyright © 2012, Hitachi, Ltd.

(10) Pair Status Operation

A Pair Status operation makes the MCU or the RCU display the result of the Add Pair operation or the Pair Status operation to the specified Remote Copy volume pair, along with the following information:

Pair Synchronized	The value indicates the percent completion of the initial copy operation. This value is always 100% after the initial copy operation is complete. For a volume being queued, “Queuing” is displayed.
Pair Status	It indicates the status of the P-VOL or the S-VOL. Definition of the volume states is described in “3.10.3(3) TrueCopy Volume Pair Status”.
Last Update	Indicates the time stamp when the volume pair status has been updated. Note that the time stamp value is obtained from an internal clock in the storage system.
Pair Established	It indicates the time stamp when the volume pair has been established by an Add Pair operation. Note that the time stamp value is obtained from an internal clock in the storage system.

(11) Resync Pair Operation

A Resync Pair operation restarts the suspended Remote Copy volume pair. It also provides function to modify the Remote Copy options which will be applied to the selected Remote Copy volume pair.

 Copyright © 2012, Hitachi, Ltd.

(12) Port Operation

(a) Fibre channel interface connection

You must set up the connection port of MCU and RCU prior to the path formation in Initiator port or the RCU target port from the usual target port.
Port topology of Initiator port and the RCU target port must be set up as follows.

- Direct connection : Fabric = OFF,FC-AL
- A connection via Switch : Fabric = ON, FC-AL or Point to point
- A connection via CN2000 : Fabric = OFF, Point to point

In case of a direct connection between MCU and RCU, each Fibre channel port topology must be the same as “Fabric:Off and FC-AL”.

In case of via FC-Switch connection between MCU and RCU, each Fibre channel port topology must be set suitable for the closest FC-Switch’s topology.
(Eg.) “Fabric:On and FC-AL” or “Fabric:On and Point-to-Point” or “Fabric:Off and Point-to-Point”

 Copyright © 2012, 2013, Hitachi, Ltd.

(13) Remote Copy function Switch Operation

The Specification of a present function switches is displayed.
Moreover, the function switches which want to be specified/released can be set.
Number of function switches is 64.
This function is only SVP operation.

The function allocated in each switch is as follows.

00 ~ 12: Reserved.
13: Unused.
14: Reserved.
15: It is effective when function switch #17 is on. And the path failure threshold values are changed by combination of function switch #15 and #20.
(Refer to 17: <The path failure threshold values>)
16: Reserved.
17: The path is blocked when the number of path failures (LIP, RSCN, Time Over) reaches the threshold within a certain period. And performance decrease of remote copy caused by using the trouble path is prevented. The path failure threshold values are changed by combination of function switch #15 and #20.
<The path failure threshold values>

Function Switch			The path failure threshold values		
Switch #17	Switch #15	Switch #20	LIP	RSCN	TOV
ON	OFF	OFF	15 times	15 times	5 times
	OFF	ON	10 times	10 times	4 times
	ON	OFF	7 times	7 times	3 times
	ON	ON	3 times	3 times	2 times

18: In the case where the switch #17 is turned on, the count of link failures that occur in the fibre path is also a factor that results in a detachment caused by an excess of a threshold value.
19: Unused.
20: It is effective when function switch #17 is on. And the path failure threshold values are changed by combination of function switch #15 and #20.
(Refer to 17: <The path failure threshold values>)
21: Reserved.
22 ~ 29: Unused.
30: The following functions are supported.
- When the Pair from Storage Navigator is formed, the pair is formed with SyncCTG#7F.
- When all the connections from MCU-RCU PATH cut, S-VOL that belongs to SyncCTG#7F is suspended.

31: Unused.
32: Reserved.

Copyright © 2013, Hitachi, Ltd.

33: In the cases that a PDCM function of McData ES3232 is being used on MCU-RCU path, when response of LOGIN response is late, path status of MCU is changed to non-normal.

34 ~ 35: Unused.

36: Reserved.

37 ~ 40: Unused.

41 ~ 42: The path is blocked when the number of RIO response time that exceeds the decided time reaches the threshold within a certain period. The decided time is changed by combination of function switch #41 and #42.

<The threshold value of RIO response time delay>

Function Switch		RIO response time that is counted	The threshold value that path is blocked
Switch #41	Switch #42		
OFF	ON	10 seconds or more	100 times/10 minutes
ON	OFF	5 seconds or more	
ON	ON	2.5 seconds or more	

43: In the function of switch #17 or switch #41 and #42, the path is blocked even if the number of path is less than that is set by minimum path. Remote copy pair is suspended when the number of path is less than that is set by minimum path. So switch #43 is set up when it gives priority to response time for host over maintenance of pair state of remote copy.

<Blockade of path less than minimum path>

Switch #43	Blockade of path
OFF	Not blocked
ON	blocked

44 ~ 63: Unused.

(14) Connection composition

Connection composition examples are shown below:

In case of a direct connection between MCU and RCU, each Fibre channel port topology must be the same as “Fabric:Off and FC-AL”.

In case of via FC-Switch connection between MCU and RCU, each Fibre channel port topology must be set suitable for the closest FC-Switch’s topology.

(Eg.) “Fabric:On and FC-AL” or “Fabric:On and Point-to-Point” or “Fabric:Off and Point-to-Point”

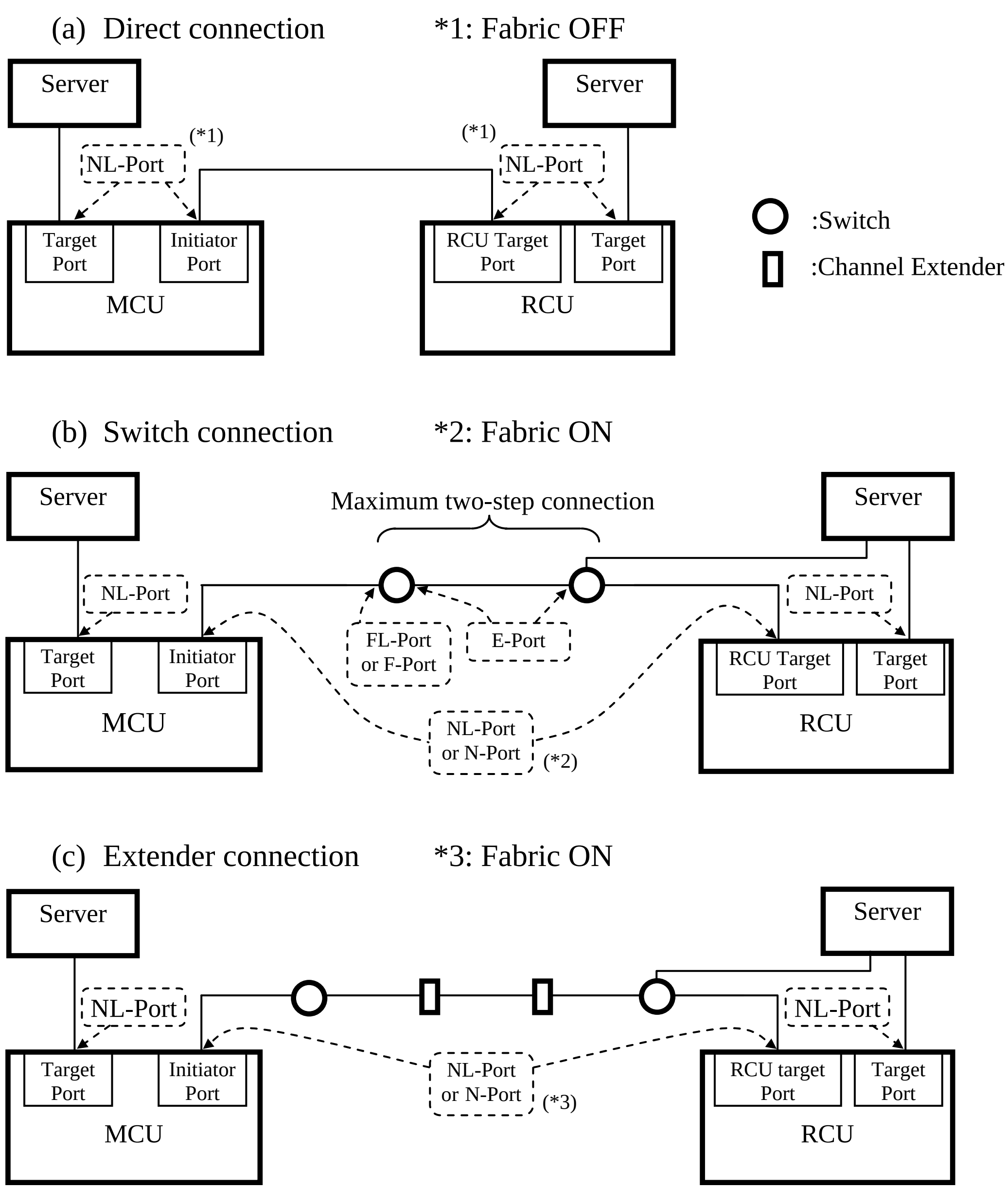


Fig. 3.10.4-2 Connection composition examples

 Copyright © 2012, Hitachi, Ltd.

3.10.5 Managing TrueCopy Environment

(1) Setting Up TrueCopy Volume Pairs

(a) Sequence of Operations
Sequence of operations to establish the TrueCopy volume pairs are shown below.

Table 3.10.5-1 Operations to Set Up TrueCopy Volume Pairs

Step		Operation	Interface
1	Set appropriate Fibre interface ports of MCU to the Initiator mode, and of RCU to the RCU Target mode.	Set Port	SVP/WebConsole
2	Establish logical paths between the TC storage systems.	Add RCU	SVP/WebConsole, RAID Manager
3	Ensure that the S-VOLs are offline from host servers.	—	—
4	Create TC voulme pairs.	Create Pair	SVP/WebConsole, RAID Manager

Several volume pairs can be specified within one Paircreate operation. After completing a Paircreate operation, another Paircreate operation can be executed to establish another TrueCopy volume pairs.

Be sure that S-VOLs are offline from host servers, before executing the create pair operation. The RCU will reject the write I/O operations to the S-VOLs once create pair operation has accepted.

Copyright © 2012, Hitachi, Ltd.

(b) Considering TrueCopy Parameters

Setting of the “fence level” parameter to the create pair operation depends on your disaster recovery planning. Refer to “3.10.6 (1) Preparing for Disaster Recovery” for these parameters.

Setting of the “minimum paths” parameter to the Add RCU operation depends on your performance requirement to the storage system at the primary site. Refer to “3.10.4 (1) Add RCU operation” for these parameters.

Setting of the “maximum initial copy activities” parameter to the System Option operation and the “priority” and the “initial copy pace” parameters to the create pair operation can control performance effect from the initial copy activities. Refer to “3.10.5 (1)(c) Controlling Initial Copy Activities” for more description.

Refer to “3.10.4(1) Add RCU operation” and “3.10.4(6) Paircreate operation” for other parameters.

(c) Controlling Initial Copy Activities

To control performance effect from the initial copy activities, the “maximum initial copy activities” parameter and the “priority” and the “copy pace” parameters can be specified:

- The “maximum initial copy activities” parameter controls the number of volumes for which the initial copy are concurrently executed;
- The “priority” parameter specifies the executing order of the initial copy on volume pair basis;
- The “copy pace” parameter specifies how many tracks should be copied by each initial copy activity.

Refer to the following example for the “maximum initial copy activities” and the “priority” parameters.

Example

Conditions:

- The create pair operation specifies that devices 00-05 should be P-VOLs.
- “Maximum initial copy activities” is set to “4” (this is the default value).
- “Priority” parameters for devices 00~05 are set to “3”, ”5”, ”5”, “1”, “4”, and “2” respectively.

Under the above conditions, the MCU will performs the initial copy:

- for devices 00, 03, 04 and 05 immediately.
- for device 01 when one of the initial copy has been terminated.
- for device 02 when the initial copy for the second device has been terminated.

(2) Suspending and Resuming the TrueCopy Volume Pairs

This section describes the operations to suspend or resume the TC volume pair, which are necessary for the following sections in this chapter.

The Suspend Pair operation with the “S-VOL” option parameters can suspend the specified TC volume pairs while the P-VOLs are still accessed from the attached host servers.

To resync the splitted TC volumes pairs, the Resync Pair operation must be executed.

Refer to “3.10.4 (8) Split Pair Operation” and “3.10.4 (6) Paircreate Operation” for more detailed description.

Copyright © 2012, Hitachi, Ltd.

(3) Managing Power On/Off of TrueCopy Components

(a) Cutting Power to TrueCopy component

Cutting power to the RCU or the Switch/Extender on the remote copy connections, or other equivalent events which make the MCU unable to communicate with the RCU should be controlled in order not to affect the remote copy activities. If the MCU detects these events when it intends to communicate with the RCU, it would suspend all TC volume pairs.

To avoid this problem, the applications on the primary host servers must be terminated or all TC volume pairs must be suspended or terminated, before performing these events.

Refer to "3.10.5(2) Suspending and Resuming the TC Volume Pairs" for the operations to suspend and resume the TC volume pairs.

(b) Power Control Interface at the Secondary Site

In the secondary site, It is not recommended to use the power control interface which remotely cuts the power to the RCU or the Switch/Extender on the remote copy connections in order to avoid the situation described in "3.10.5(3)(a) Cutting Power to TrueCopy components".

(c) Power-on-sequence

The RCU and the Switch/Extender on the remote copy connections must become operable before the MCU accepts to first write I/O operation to the P-VOLs.

After the power-on-reset sequence of the MCU, it communicates with the RCU in order to confirm the status of the S-VOLs. If it is not possible, the MCU retries the confirmation until it is successfully completed or the MCU accepts the first write I/O operations to the P-VOLs.

If the MCU accepts the first write I/O operation before completing the confirmation, the MCU will suspend the TC volume pair. This situation is critical because the status of the S-VOL can not be changed, that is, remains "PAIR" state.

3.10.6 TrueCopy Error Recovery

(1) Preparing for Disaster Recovery

(a) Considering Fence Level Parameter

Table 3.10.6-1 shows how the fence level parameter of the Paircreate operation has an effect on the write I/O operations to the P-VOL after the TC volume pair has been suspended. You should select one of the fence level considering the “degree of the currency” of the S-VOL required by your disaster recovery planning. The SVP or Web Console, which is connected to either the MCU or the RCU, can display the fence level parameter which has been set to the TC volume pairs.

Table 3.10.6-1 Effect of the Fence Level Parameter

Failure		Subsequent write I/O operations to the P-VOL will be ...		
		“Data”	“Status”	“Never”
1)	The update copy has failed,	Rejected	—	—
2)	(1) & however the status of the S-VOL could have been successfully changed to “PSUS/PSUE” state.	Rejected	accepted	accepted
3)	(1) & furthermore the status of the S-VOL could not have been changed to “PSUS/PSUE” state.	Rejected	Rejected	accepted

NOTE: “Data” and “Status” has an effect when an TC volume pair of “PAIR” state is suspended. For TC volume pairs which are in “COPY” state, subsequent write I/O operations will not be rejected regardless of Fence Level parameter.

 Copyright © 2012, Hitachi, Ltd.

1) Fence Level = “Data”

The data of the S-VOL is always identical to the P-VOL if once the TC volume pair has been successfully synchronized. You can reduce the time to analyze whether the S-VOL is current or not in your disaster recovery procedures.

However, this parameter will make the P-VOL not accessible from your applications whenever the TC copy activity has failed. Therefore you should specify this parameter to the most critical volumes for your disaster recovery planning.

Most of the database system supports duplexing the critical files, for example log files of DB2, for its file recovering capability. It is recommended to locate the duplexed files on the volumes in the physically separated storage systems, and establish TC volume pairs for each volumes by using physically separated remote copy connections.

NOTE: If the failure has occurred before completing the initial copy, the S-VOL can not be used for disaster recovery because the data of the S-VOL is not fully consistent yet. You can became aware of this situation with referring status of the S-VOL in your disaster recovery procedures. Refer to “3.10.6(2)(b) Analyzing the Currency of S-VOLs” for more detailed description.

NOTE: Only the difference between the TC volume pair must be the last update from the host server. TC is a synchronous remote copy. The MCU reports a “unit check” if it detects the failure on the write I/O operation including the update copy to the S-VOL. Therefore, the operating system and the application program does not regard the last (failed) I/O operation as successfully completed.

Copyright © 2012, Hitachi, Ltd.

2) Fence Level = “Never”

The subsequent write I/O operations to the P-VOL will be accepted even if the TC volume pair has been suspended. Therefore the contents of the S-VOL can become “older” (behind the currency of corresponding P-VOL) if the application program continue updating the P-VOL. Furthermore, the status of the S-VOL which will be obtained from the RCU can not be in a “PSUS/PSUE” state.

To use this parameter, your disaster recovery planning must satisfy the following requirements:
- The currency of the S-VOL should be decided by analyzing the S-VOL itself with other files which are confirmed to be current.
- The data of the S-VOL should be recovered by using other files which are ensured to be current.

3) Fence Level = “Status”

The level of this parameter is between “Data” and “Never”. Only when the status of the S-VOL can be ensured, the subsequent write I/O operations to the P-VOL will be permitted. Therefore the disaster recovery procedure of deciding the currency of the S-VOL can be reduced.

Copyright © 2012, Hitachi, Ltd.

(b) File Recovery Procedures Depending on Installations

TC is a synchronous remote copy. All updates to the P-VOLs are copied to their S-VOLs before completing each channel program of the write I/O operations. When the TC volume pairs have been suspended or the MCU has become inoperable due to a disaster, therefore, many data “in progress” could remain in the S-VOLs. That is, some data set might be still opened, or some transactions might not be committed yet. All breakdown cases should be previously considered.

Therefore, even if you have selected the fence level of “Data” for all TC volume pairs, you should establish the file oS-VOLume recovery procedures. The situation which should be assumed is similar to that where the volumes have became not accessible due to the disk controller failure in non-remote copied environment.

If you use the fence level of “Status” or “Never”, the suspended S-VOLs could become “ancient” compared to otheS-VOLumes. This situation might cause a data inconsistency problem among several volumes.
You should prepare, in your disaster recovery, for recovering some files or some volumes which have become “ancient” by using:
- files for file recovery, for example DB2 log files, which have been confirmed to be current. To ensure the currency of these files, it is recommended to use the fence level of “Data” for these critical volumes.
- the sense information with the system time stamp which have been transferred through the error reporting communications.
- full consistent file oS-VOLume backups, if the sense information and the system time stamp can not be used.

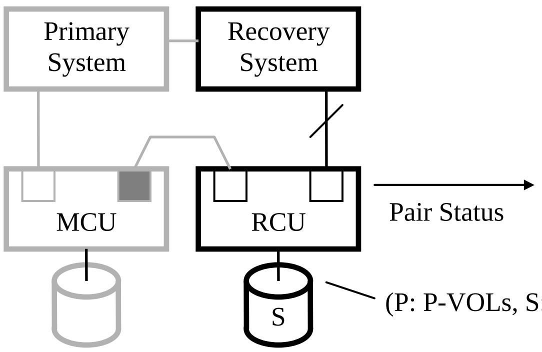
 Copyright © 2012, Hitachi, Ltd.

(2) Disaster Recovery Procedures - Switching to the Recovery System

(a) Summary

① Primary system and MCU becomes inoperable due to disaster.

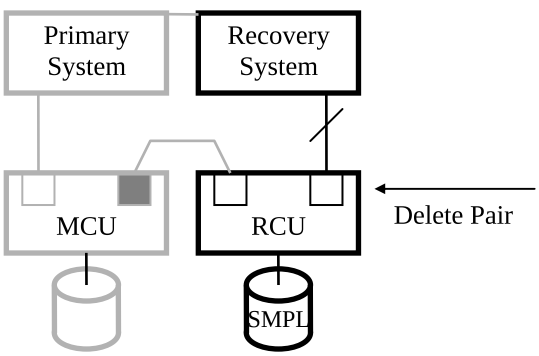
②



(P: P-VOLs, S: S-VOLs, SMPL: Simplex Volumes)

Step1: Analyze currency of the S-VOLs.
- Pair Status, Fence Level from the RCU

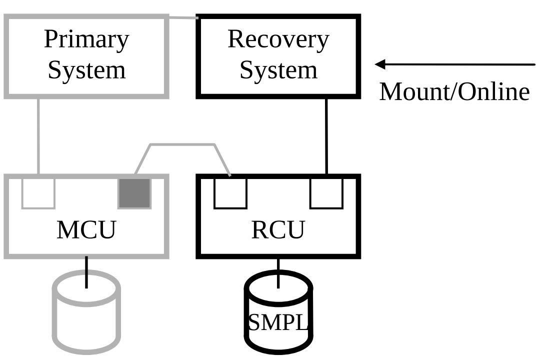
③



Step2: Terminate all TC volume pairs.
- Delete Pair to the RCU
- S-VOLs are changed to “SMPL” .

➔ Refer to 3.10.6(2)(c)

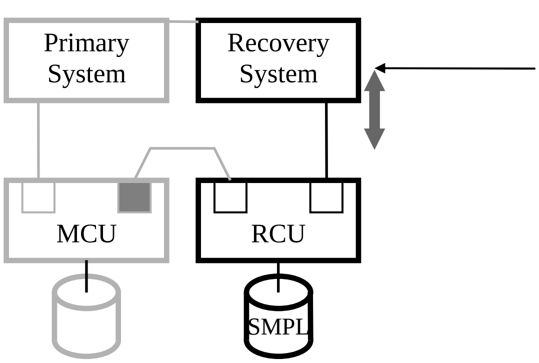
④



Step3: Mount/Online all S-VOLs.
- Some volumes may require file recovery before being brought online.

➔ Refer to 3.10.6(2)(d)

⑤



Step4: Make the S-VOLs current and restart application recovery procedure.

➔ Refer to 3.10.6(1)(c)

Copyright © 2012, Hitachi, Ltd.

(b) Analyzing the Currency of S-VOLs (Step 1)

1) Analyzing Status of the S-VOLs and Fence Level Parameter

Table 3.10.6-2 Currency of the S-VOLs

	Fence Level for this TC volume pair		
Status of S-VOL	Data	Status	Never
SMPL	To be confirmed	To be confirmed	To be confirmed
COPY	Inconsistent	Inconsistent	Inconsistent
PAIR	Current	Current	To be analyzed
PSUS/PSUE (Initial Copy Failed)	Inconsistent	Inconsistent	Inconsistent
PSUS/PSUE (by other reason)	Current	Suspected	Suspected

Table 3.10.6-2 shows how to analyze the currency of the S-VOLs referring the status of the S-VOLs and the fence level parameter which have been specified when establishing the TC volume pairs.
The status of the S-VOLs must be obtained from the RCU in your disaster recovery procedures.
The fence level parameter must be previously field since it cannot be obtained From RCU.

The meaning of the results or further actions shown in each column of Table 3.10.6-2 are as follows:

To be confirmed This volume does not belong to any TC volume pair. If you have certainly established the TC volume pair for this volume and you have never deleted it, you should regard this volume as inconsistent.

Inconsistent The data on this volume is inconsistent because not all data of P-VOL have successfully been copied to this volume yet. You can not use this volume for the applications unless this volume is initialized (or successfully copied from the P-VOL at later time).

Current The data on this volume is completely synchronized with the corresponding P-VOL.

To be analyzed The currency on this volume can not determined. To determine the currency, further analysis described in (2) of this section should be performed.

Copyright © 2012, Hitachi, Ltd.

Suspected The data on this volume must be "older", behind the currency of corresponding P-VOL. You should restore the consistency of this volume at least, and the currency of this volume if required. Time of suspension obtained from the Pair Status operation will help you decide the last time when this volume was current.

2) Further Analysis by Referring to Other Information

The P-VOLs, to which the fence level parameter has been set to "Never", will accept the subsequent write I/O operations regardless of the result of communication to change the S-VOL into the "PSUS/PSUE" state. Therefore, the status of the S-VOL should be analyzed by referring to the following information:

- The status of the P-VOL obtained from the MCU, if possible. You should return to Table 3.10.6-2 with assumption of the same status as the P-VOL and fence level of "Status".
- The other related files, for example DB2 log files, which have been confirmed to be current.

(c) Terminating TrueCopy Volume Pairs (Step 2)

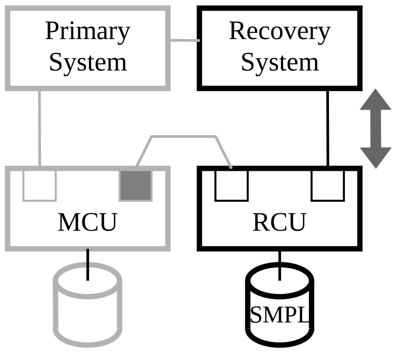
The "Delete Pair" operation to the RCU terminates the specified TC volume pairs. These S-VOLs will be changed to "SMPL" state.

Copyright © 2012, Hitachi, Ltd.

(3) Disaster Recovery Procedures - Returning to the Primary Site

(a) Summary

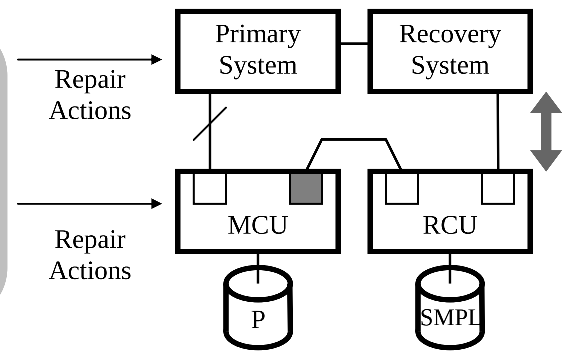
①



Applications are working at the recovery system.

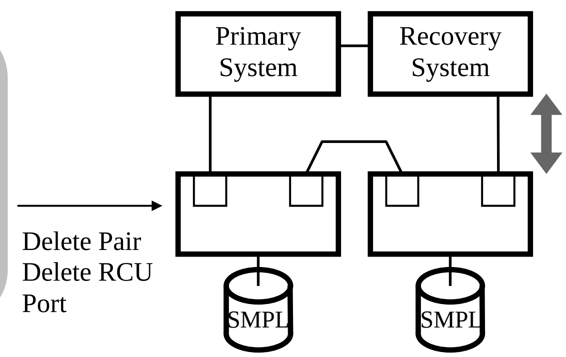
②

Step1: Make the components in the primary system operable.

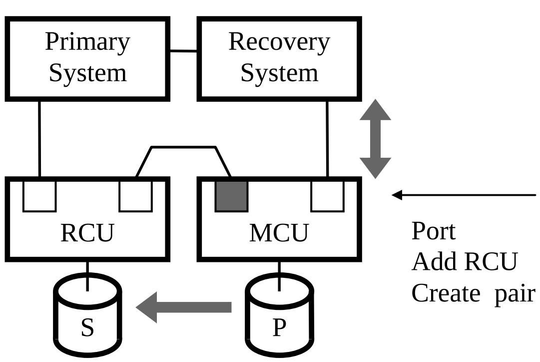


③

Step2: Terminate the TC settings remaining in the MCU.
(1) Delete Pair: Delete all TC volume pairs.
(2) Delete RCU
(3) Port: Change from Initiator to RCU Target mode.
➔ Refer to 3.10.6(3)(b)



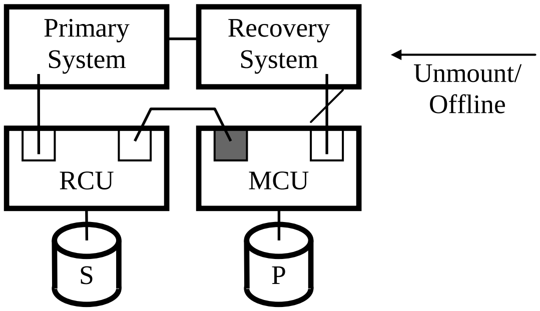
④



Step3: Create TC with the reverse direction
(1) Port: Change from RCU Target to Initiator mode.
(2) Add RCU
(3) Create Pair
➔ Refer to 3.10.6(3)(c)

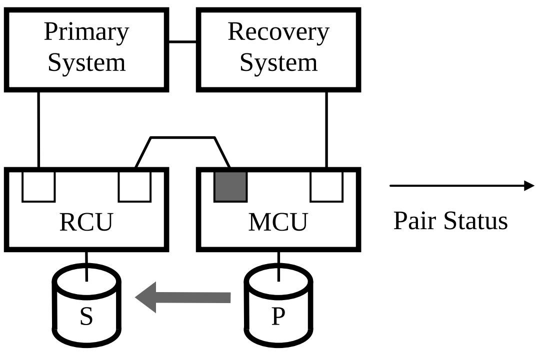
Copyright © 2012, Hitachi, Ltd.

⑤



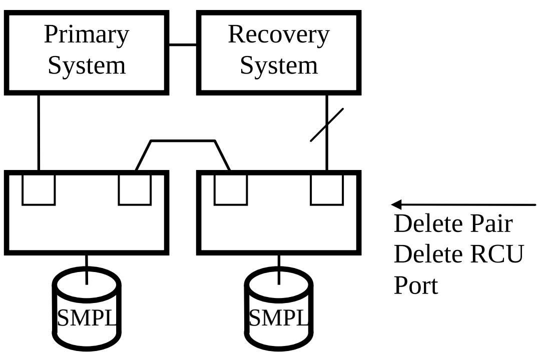
Step4: Halt related applications and vary all P-VOLs offline from the recovery system.

⑥



Step5: Confirm all TC volume pairs become “PAIR” state.

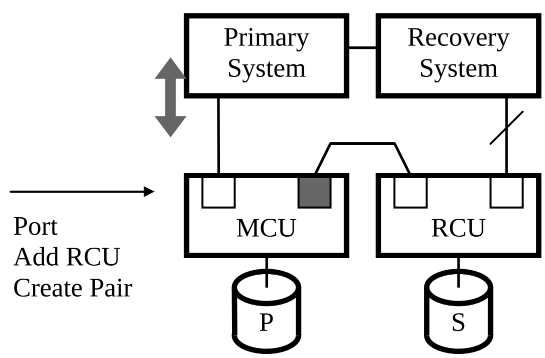
⑦



Step6: Terminate all TC settings.
(1) Delete Pair: Delete all TC volume pairs.
(2) Delete RCU
(3) Port: Change from Initiator to RCU Target mode.
➔ Refer to 3.10.6(3)(d)

⑧

Step7: Create TC pair with the original direction and start applications.
(1) Port: Change from RCU Target to Initiator mode.
(2) Add RCU
(3) Create Pair
➔ Refer to 3.10.6(3)(e)



(b) Terminating the TrueCopy Settings Remaining in the MCU (Step2)

After the storage system becomes operable, the remaining registration of the TC volume pairs and the RCU should be deleted by performing the Delete Pair operation and Delete RCU operation respectively.

Note that the status of P-VOLs may be “PSUS (Delete Pair to RCU)” because of Delete Pair operation issued to the RCU in step2 of “3.10.6(2) Disaster Recovery Procedures - Switching to the Recovery System”. It is normal condition in this situation.

Before performing the Delete RCU operation, all TC volume pairs must be deleted.

If you want to use same remote copy connections for step 3, the fibre interface ports which have been set to the Initiator mode should be changed to the RCU Target mode by the Port operation.

(c) Establish TrueCopy with the Reverse Direction (Step3)

The TC volume pair should be established with the reverse direction to synchronize the original P-VOLs with the original S-VOLs. The procedures for this step are same as those described in “3.10.5(1) Setting Up TC Volume Pairs”. Note that the storage systems in the original primary site and the recovery site are treated as the RCUs/S-VOLs and the MCUs/P-VOLs respectively.

Do not select “none” parameter to the Paircreate operations. The volumes in the original primary site are now behind the volumes in the recovery site.

Copyright © 2012, Hitachi, Ltd.

(d) Terminate Applications and TrueCopy Settings at the Recovery Site (Step 4 ~ 6)

TC settings with the reverse direction must be deleted after halting the applications in the recovery site (step 4) and confirming that all TC volume pairs are in “PAIR” state (step 5).

If you want to use same remote copy connections for step 7, the fibre interface ports which have been set to the Initiator mode should be changed to the RCU Target mode by the Port operation.

(e) Create TrueCopy Pair with the Original Direction and Start Applications (Step 7)

The TC volume pair should be established with the original direction to synchronize the original P-VOLs with the original S-VOLs. The procedures for this step are same as those described in “3.10.5(1) Setting Up TrueCopy Volume Pairs”.

Do not select “none” parameter to the Paircreate operations. The volumes in the original primary site are now behind the volumes in the recovery site.

Copyright © 2012, Hitachi, Ltd.

3.11 ShadowImage

3.11.1 Overview

(1) Main object

1) Reduce Backup time.
2) Easy testing before system upgrade with the data whose applications are actually used on the system.

(2) Function Outline

1) Making duplicated volumes.
2) There is no conflict on volume because the duplicated volumes are on another physical storage system.
3) Three destination volumes can be with one master volume.
   Those three pairs can be split independently.
4) ShadowImage can be controlled from RAID Manager.

Copyright © 2012, Hitachi, Ltd.

3.11.1.1 Outline of ShadowImage

No.	Items	Specification
1	Coupling object	One logical volume (LDEV)
2	Requirement for create a pair	(1) Pair LDEVs have to be a same track format and same capacity. (2) Pair LDEVs have to exist in a same storage system. (3) It is not possible to share a destination volume at same time.
3	Support of Customized Volume Size (CVS)	SI : Supported
4	Combination of RAID level between master volume and destination volume	RAID1(2D+2D)←→RAID1(2D+2D) RAID5(3D+1P or 7D+1P)←→RAID5(3D+1P or 7D+1P) RAID5(3D+1P or 7D+1P)←→RAID1(2D+2D) RAID6(6D+2P)←→RAID6(6D+2P) RAID6(6D+2P)←→RAID1(2D+2D) RAID6(6D+2P)←→RAID5(3D+1P or 7D+1P)
5	Data protection	There is a parity protection for both master volume and destination volume.
6	RESYNC pattern	SI supports 2 types of RESYNC pattern. From Master Volume data to Destination volume and from Destination Volume to Master Volume
7	Time for transition from Duplex to Split.	3 min./VOL (3390-3) without I/O (Depend on the number of pairs and the load of DKC)
8	When the destination volume can be accessed from HOST.	The destination volume can be accessed at only Split status.

(To be continued)

Copyright © 2012, Hitachi, Ltd.

(Continued from the preceding page)

No.	Items	Specification
9	Cooperation with TrueCopy	SI : Supported The master volume of SI can be an M-VOL or R-VOL of TC. • TC volumes are shared with Universal Replicator volumes can't be SI pair volumes.
	Cooperation with TrueCopy (only for ShadowImage)	Supported • The M-VOL or R-VOL of TC can be a primary volume of SI. • The secondary volume of SI can be an M-VOL of TC. • TC volumes are shared with Universal Replicator volumes can be SI pair volumes.
10	Cooperation with Volume Migration	Supported • The source, destination or RESERVE volume of Volume Migration can not be the primary, secondary or RESERVE volume of SI. • The primary, secondary or RESERVE volume of SI can be the source volume of Volume Migration. But if SI P-VOL or RootVOL already has 3 pairs, it can not be the source volume of Volume Migration. • When SI pair which is combined with Volume Migration is split, the migration of Volume Migration is canceled.
11	Cooperation with Universal Replicator (only for ShadowImage)	Supported • Universal Replicator volumes (primary / secondary) can be a primary volume of SI. • The secondary volume of SI can be Universal Replicator volumes (primary / secondary). • The journal volume of Universal Replicator can't be a primary or secondary volume of SI. • Universal Replicator volumes are shared with TC volumes can be SI pair volumes.

(To be continued)

Copyright © 2012, Hitachi, Ltd.

(Continued from the preceding page)

No.	Items	Specification
12	At-Time Split Function (ShadowImage)	The SI At-Time Split function applies to SI pairs that belong to a consistency group, and enables you to create Source volumes of all Target volumes in the same consistency group, at the time when the pairsplit command is executed using the Command Control Interface (CCI) software from the UNIX®/PC server host to the 9900V storage system. NOTE: For further information on Command Control Interface, please refer to “Hitachi Command Control Interface User and Reference Guide”. An SI consistency group is a user-defined set of SI volume pairs used for the At-Time Split function. Users can defined a consistency group by using CCI on the UNIX®/PC server host. SI consistency groups also correspond to the groups registered in the CCI configuration definition file. SI consistency groups have the following restrictions: • You can configure up to 128 consistency groups in a storage system. • A number (0~7F) is assigned to each consistency group. You can specify a consistency group number when you create SI pairs. If you do not specify a number, then the 9900V storage system assigns a number automatically. • You can define up to 8,192 SI pairs in a consistency group. However, for LUSE volumes that contain n LDEVs, you should count as n pairs. NOTE: For further information on LUSE volumes, please refer to “Provisioning Guide”. • To configure SI consistency groups, you can only use the CCI software. However, to confirm the SI consistency group numbers, you can also use the Storage Navigator.
13	The maximum number of pairs	The maximum number of pairs is as follows. The maximum number of pairs is the sum of the numbers of pairs of the SI and Volume Migration.
14	The Thin Image Optional function	The Thin Image Optional function uses the Virtual Volume (V-VOL) that has no actual volume capacity as the Secondary Volume (S-VOL). When the host access to a V-VOL, the access goes to either the Pool Volume (pool-VOL) or the Primary Volume (P-VOL) depending on whether the area on the P-VOL has been updated or not. The Pool Volume stores the before-image of data on P-VOL, which is copied to the Pool Volume before the host updates P-VOL. Each address on V-VOL is mapped to P-VOL or pool-VOL, and the mapping information (The Virtual Volume Mapping Information) is stored in the Shared Memory and the Pool Volume.

Copyright © 2012, Hitachi, Ltd.

Blank Sheet

(1) Calculating Maximum Number of Pairs

When you create ShadowImage pairs, resources called differential tables, pair tables will be required. The number of available differential tables, pair tables in one storage system depends on whether the additional shared memory is installed or not. There are several patterns of additional shared memory for differential tables, pair tables, and you can choose any pattern you like. Table 3.11.1.1-1 shows the pattern of additional shared memory.

Table 3.11.1.1-1 Additional Shared Memory for Differential Tables, Pair Tables

Additional Shared Memory for Business Copy	Number of Differential Tables	Number of Pair Tables	Number of System Volumes
Base (No additional shared memory)	26,176	8,192	8,192
Extension 1	104,768	16,384	8,192
Extension 2	209,600	16,384	8,192

NOTE:

- To install additional shared memory for differential tables, please call the Support Center.
- Even if you install additional shared memory for differential tables, pair tables, the maximum number of pairs is half of the total number of volumes in the storage system (see Table 3.11.1.1-1). For example, in case of “Base (No additional shared memory)” in Table 3.11.1.1-1, when P VOLs and S VOLs are in a one-to-one relationship, you can create up to 8,192 pairs. However, note that in case of “Extension 1” and “Extension 2”, the maximum number of pairs is 16,384 regardless of the total number of volumes in the storage system.

To calculate the maximum number of ShadowImage pairs, first you need to calculate how many differential tables, pair tables are required to create ShadowImage pairs, and then compare the result with the number of differential tables, pair tables in the whole storage system. Note that in addition to ShadowImage, the following program products will also use differential tables, pair tables.

The following program products use differential tables
- Volume Migration

The following program products use pair tables
- Volume Migration

Copyright © 2012, Hitachi, Ltd.

If ShadowImage and these program products are used in the same storage system, the number that is deducted the number of differential tables, pair tables used by the pairs (migration plans in case of Volume Migration) of the program products shown in above from the number of differential tables, pair tables in the whole storage system will be the number of available differential tables, pair tables for ShadowImage pairs.

Please refer to "Volume Migration User Guide" to calculate the number of differential tables, pair tables that are required for Volume Migration.

Assuming that only ShadowImage uses differential tables, pair tables, this section describes how to calculate the number of differential tables, pair tables required for one ShadowImage pair, and the conditions you need to consider when calculating the number of ShadowImage pairs that can be created.

NOTE: You can use CCI's inqraid command to query the number of the differential tables required when you create ShadowImage pairs. You can also query the number of the differential tables not used in the storage system by using this command. For details about inqraid command, please refer to "Hitachi Command Control Interface User and Reference Guide".

Copyright © 2012, Hitachi, Ltd.

(1-1) Calculation of the Number of Differential Tables, Pair Tables Required for One Pair

When you create a ShadowImage pair, the number of required differential tables, pair tables will change according to size of the volumes. To calculate the number of differential tables, pair tables required for a pair according to size of the volumes, use the expression in Table 3.11.1.1-2.

Total number of the differential tables per pair = ((X) ÷ 256) ÷ (Z)

(X): The capacity of the volume. (KB) (*1)

(Z): 20,448 (The number of the slots that can be managed by a differential table)

Note that you should round up the number to the nearest whole number. For example, when provided that the number of the cylinders is 3,019,898,880 KB ((X) in the expression above), the calculation of the total number of the differential tables is as follows.
(3,019,898,880 ÷ 256) ÷ 20,448 = 576.9014...
When you round up 576.9014 to the nearest whole number, it becomes 577. Therefore, the total number of the differential tables for one pair is 577.
In addition, 1 pair tables per 36 differential tables is used. The number of pair tables used for above- mentioned 17.

*1: If the volume is divided by VLL function, you need to apply the capacity of the volume after the division.

 Copyright © 2012, Hitachi, Ltd.

(1-2) Conditions for the Number of ShadowImage Pairs that can be Created (In Case of No LUSE Volume Exists)

The number of ShadowImage pairs that can be created will change according to whether LUSE volumes are used or not. This section describes the how to calculate the number of ShadowImage pairs that can be created, when no LUSE volume is used (e.g. 1 LU consists of one volume).

When you do not use LUSE volume, you can use the following inequation to know whether you will be able to create the desired number of ShadowImage pairs or not.

NOTICE: You need to meet the inequation below:
Σ {(α) × (the number of ShadowImage pairs)} ≤ (β)
and
Σ {(γ) × (the number of ShadowImage pairs)} ≤ (δ)

- (α): The required number of differential tables per pair.
  (α) changes according to size of the volume. For information about how to calculate (α), see "(1-1) Calculation of the Number of Differential Tables, Pair Tables Required for One Pair" (THEORY03-11-80).
- (β): The number of differential tables available in the storage system.
  If an additional shared memory for differential tables is not installed, (β) is 26,176. If an additional shared memory for differential tables is installed, see Table 3.11.1.1-1.
- (γ): The required number of pair tables per pair.
  Value of (γ) changes according to the (α). For information about how to calculate (γ), see "(1-1) Calculation of the Number of Differential Tables, Pair Tables Required for One Pair" (THEORY03-11-80).
- (δ): The number of pair tables available in the storage system.
  If an additional shared memory for pair tables is not installed, (δ) is 8,192. If an additional shared memory for pair tables is installed, see Table 3.11.1.1-1.

For example, if you are to create 20 pairs of in a storage system that is not installed with an additional shared memory for differential tables, pair tables, you can use the condition inequation as follows:

When the capacity of the volume is 3,019,898,880 KB, the number of differential tables required for a pair will be 577. The number of pair tables required for a pair will be 17.
If you apply these numbers to the above-mentioned inequation:
$3 \times 10 + 577 \times 20 = 11{,}570 \leq 26{,}167$
and
$1 \times 10 + 17 \times 20 = 350 \leq 8{,}192$
Since 11,570 is smaller than 26,167, you can see that 20 pairs can be created.

Copyright © 2012, Hitachi, Ltd.

(1-3) Conditions for the Number of ShadowImage Pairs that can be Created (In Case of LUSE Volume Exists)

The number of ShadowImage pairs that can be created will change according to whether LUSE volumes are used or not. This section describes how to calculate the number of ShadowImage pairs that can be created, when LUSE volumes are used.
When you use LUSE volume, you can use the following inequation to know whether you will be able to create the desired number of ShadowImage pairs or not.

NOTICE: You need to meet the inequation below.
Σ[Σ{(α) × (the number of the volumes that forms LUSE volumes)} × (the number of ShadowImage pairs)] ≤ (β)
and
Σ[Σ{(γ) × (the number of the volumes that forms LUSE volumes)} × (the number of ShadowImage pairs)] ≤ (δ)

- (α): The required number of differential tables for each volume that forms LUSE volume.
  When you use LUSE volumes to create a ShadowImage pair, every volume that forms the LUSE volume will use the differential tables as a pair. For example, if you create a ShadowImage pair by using LUSE volumes which are created by combining two volumes, you need differential tables for two pairs.
  (α) changes according to size of the volume. For information about how to calculate (α), see “(1-1) Calculation of the Number of Differential Tables, Pair Tables Required for One Pair” (THEORY03-11-80).
- (β): The number of differential tables available in the storage system.
  If an additional shared memory for differential tables is not installed, (β) is 26,176. If an additional shared memory for differential tables is installed, see Table 3.11.1.1-1.
- (γ): The required number of pair tables for each volume that forms LUSE volume.
  Value of (γ) changes according to the (α). For information about how to calculate (γ), see “(1-1) Calculation of the Number of Differential Tables, Pair Tables Required for One Pair” (THEORY03-11-80).
- (δ): The number of pair tables available in the storage system.
  If an additional shared memory for pair tables is not installed, (δ) is 8,192. If an additional shared memory for pair tables is installed, see Table 3.11.1.1-1.

For example, if you are to create 10 pairs of LUSE volumes consisting respectively of three volumes, you can use the condition inequation as follows:
When the capacity of the volume is 3,019,898,880 KB, the number of differential tables required for a pair will be 577. The number of pair tables required for a pair will be 17.
If you apply this number to the above-mentioned inequation:
$(577 \times 3) \times 10 = 17{,}310 \leq 26{,}176$
and
$(17 \times 3) \times 10 = 510 \leq 8{,}192$
Therefore, you can see that 10 ShadowImage pairs which are formed by three volumes can be created.

Copyright © 2012, Hitachi, Ltd.

In case of ShadowImage volume is selected for Volume Migration volume

	P-VOL	S-VOL	RootVOL	NodeVOL	LeafVOL	ReserveVOL
Source VOL	Possible (*1)	Possible	Possible (*1)	Possible (*2)	Command Reject	Possible
Target VOL	Command Reject	Command Reject	Command Reject	Command Reject	Command Reject	Command Reject
ReserveVOL	Command Reject	Command Reject	Command Reject	Command Reject	Command Reject	Command Reject

*1: It is impossible if ShadowImage P-VOL or RootVOL already has 3 pairs.

*2: It is impossible if ShadowImage NodeVOL is already paired with 2 LeafVOLs.

If you want to execute migration of Volume Migration, you need to delete the pair and reset ReserveVOL of ShadowImage.

Copyright © 2012, Hitachi, Ltd.

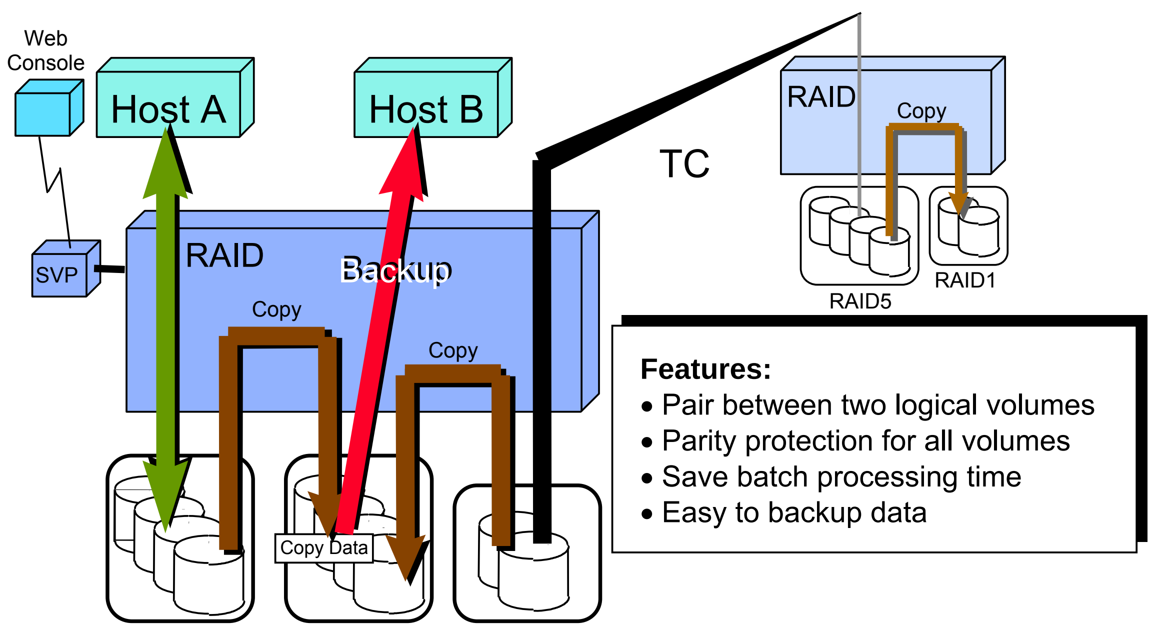
In case of Volume Migration volume is selected for ShadowImage volume

	Source VOL	Target VOL	ReserveVOL
P-VOL	Command Reject	Command Reject	Command Reject
S-VOL	Command Reject	Command Reject	Command Reject
RootVOL	Command Reject	Command Reject	Command Reject
NodeVOL	Command Reject	Command Reject	Command Reject
LeafVOL	Command Reject	Command Reject	Command Reject
ReserveVOL	Command Reject	Command Reject	Command Reject

If you want to add pair of ShadowImage, you need to cancel Volume Migration migration.

CAUTION

Copy process is done asynchronously with HOST I/O according to differential bit map.
Differential bit map is recorded on shared memory. So if shared memory is lost by offline micro exchange or volatile PSon etc., DKC lost differential bit map.
In these cases DKC treat as whole volume area has differential data, so copy process will take longer time than usual. And if the pair is SPLIT-PEND status, the pair become SUSPEND status because lost of differential bit map.
Primary volumes and secondary volumes of SI pairs should be placed on many RAID groups separately. And SI pairs which are operated at the same time should be placed in other RAID groups. SI pairs which are concentrated at very few RAID groups may influence HOST I/O performance.
If DKC is busy, increase Cache, DKB and RAID groups. And secondary volumes of SI pairs should be placed in the increased RAID groups. SI pairs in very busy DKC may influence HOST I/O performance.



3.11.2 Construction of ShadowImage

- ShadowImage can be controlled from SVP, Web Console and HOST.
- DKB with LA exchange has to exist in the storage system.

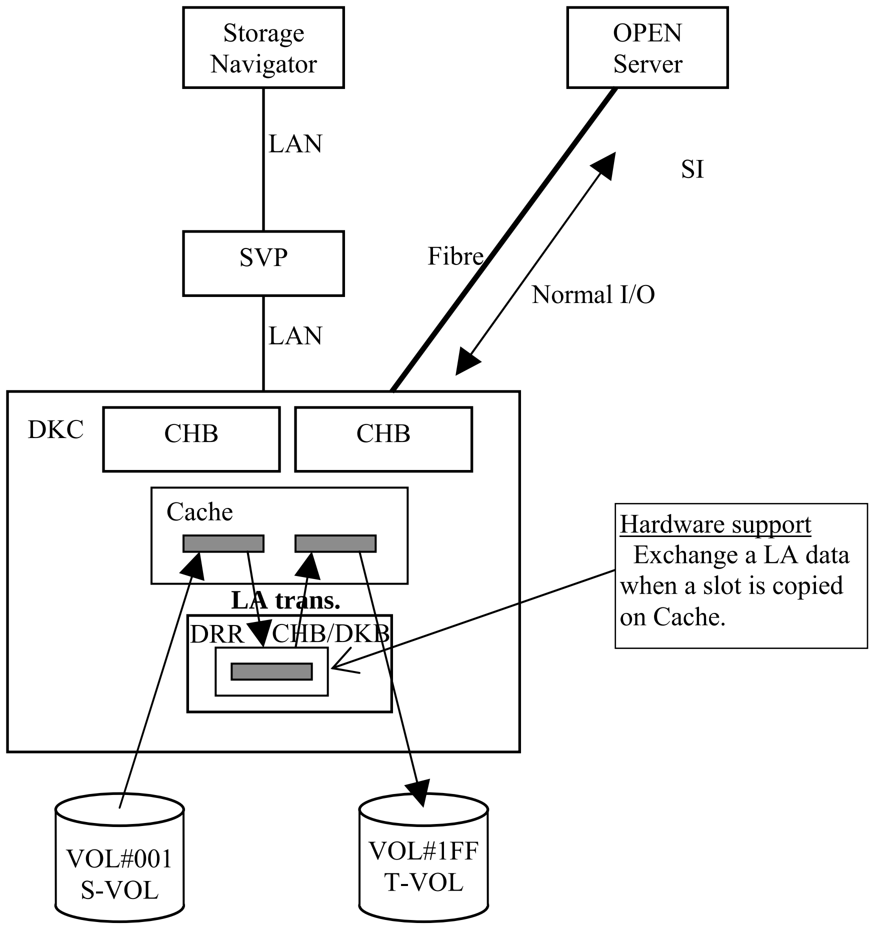


Fig. 3.11.2-1 Construction of ShadowImage

Copyright © 2012, Hitachi, Ltd.

3.11.3 Status transition

Table 3.11.3-1 Status of ShadowImage

No.	Status	Definition
1	Simplex	There is no pair with the volume.
2	Pending	In the copy job status from the master volume to the destination volume for duplex status.
3	Duplex	The copy from master to destination is finished. The destination volume can not be accessed from HOST.
4	Split Pending	In the copy job status of the differential data from the master volume.
5	Split	The pair is split. The destination volume can be accessed from HOST. In this status, the position of write data from the HOST is recorded on a bitmap to reduce the copy time on RESYNC.
6	Resync	In the copy job status of the differential data from master to destination.
7	Suspend	There is an error with the pair. After a running copy job was stopped by the SVP operation, the pair status is "suspend".

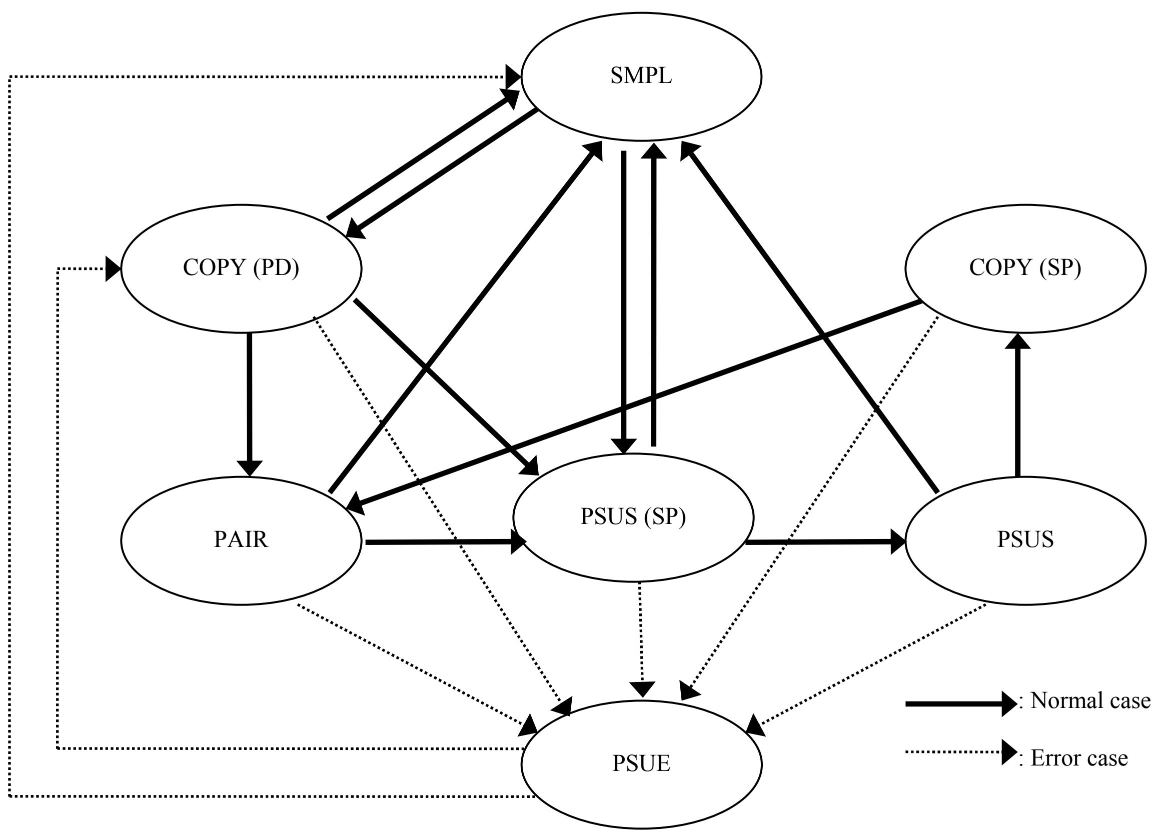


Fig. 3.11.3-1 State transition of ShadowImage

3.11.4 Interface

(1) Outline

ShadowImage support a command set to control ShadowImage functions.
This command set is a common interface in a storage system. So the commands from different HOSTs are translated to the ShadowImage command at each command process.

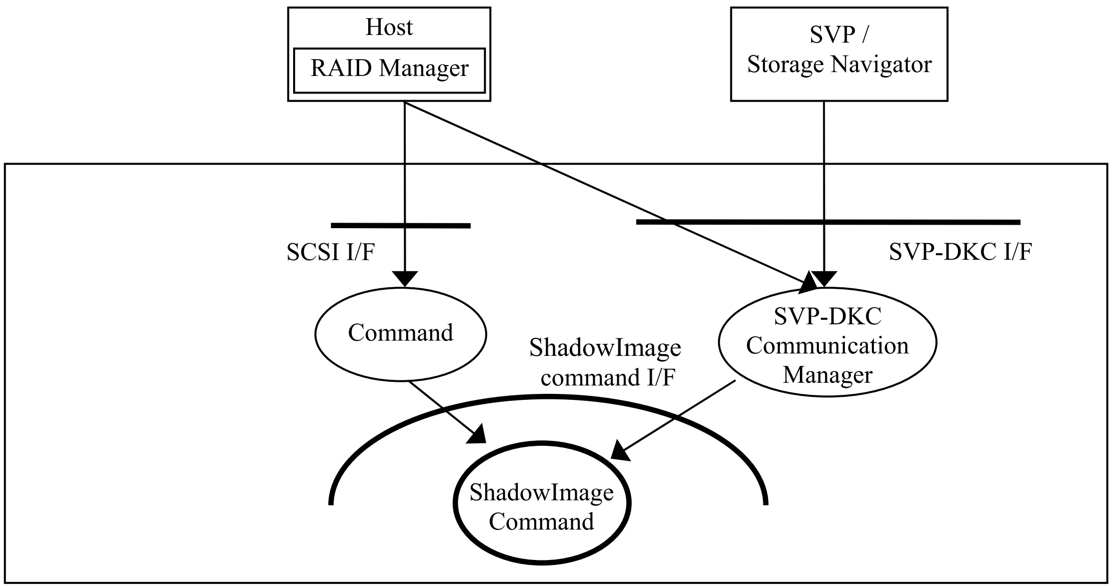


Fig. 3.11.4-1 Outline of ShadowImage IF

NOTE: It is necessary to define Command Device before using by In-band system RAID Manager on HOST.
Do not define Command Device on a heavy-load path.

(2) ShadowImage operation

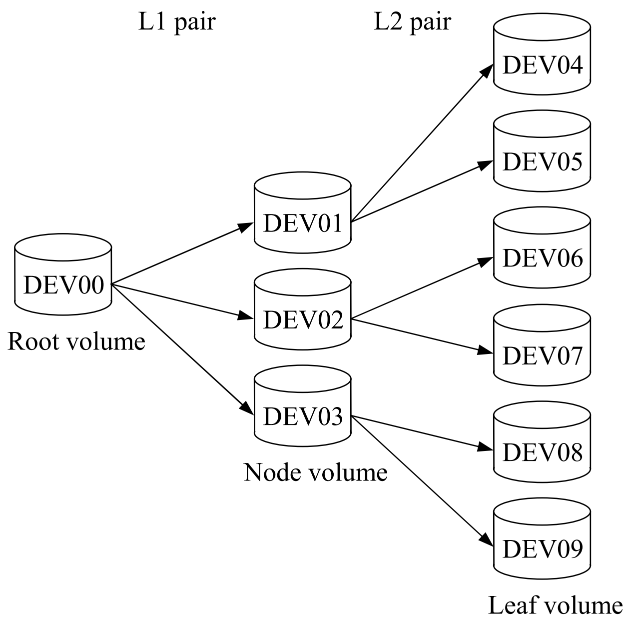
Table 3.11.4-1 ShadowImage operation

No.	Command	Operation
1	Duplex	Creates a pair and start initial copy
2	Split	Splits the pair
3	RESYNC	Resumes the pair and start differential copy
4	Simplex	Deletes the pair
5	Suspend	Suspends the pair action
6	Reserve	Marks and Unmarks the volume for a candidate of destination volume

Copyright © 2012, Hitachi, Ltd.

3.11.5 Cascade function

Cascade function makes a pair with an existed Target volume as a new Source volume. See the figure below.



No.	Content	Specification
1	Pair structure	A Target volume of L1 pair (=Node volume) can be a Source volume of L2 pair.
2	Number of copies	Root : Node = 1 : 3 Node : Leaf = 1 : 2 Root : (Node + Leaf) = 1 : 9
3	Split pair condition	L2 pair is able to execute split pair request only when the L1 pair is in the split status.
4	Delete pair condition	• No conditions. • When L1 pair is deleted, L2 pair becomes L1 pair.
5	Combination with TrueCopy	Possible. However, Node volume and Leaf volume are treated as a target volume.
6	Combination with Volume Migration	Possible. However, Leaf volume cannot be moved.

Copyright © 2012, Hitachi, Ltd.

- Name of volume type
  Source volume of the first pair : Root Volume
  Target volume of the first pair : Node Volume
  Source volume of the 2nd pair : Node Volume
  Target volume of the 2nd pair : Leaf Volume
- Name of pair
  - The first pair (A pair of root volume is source volume) : L1 pair
  - The second pair (A pair of node volume is source volume) : L2 pair
- Name of pair chain
  - A chain of L1 pair and L2 pair with a node volume: stream

Copyright © 2012, Hitachi, Ltd.

3.11.6 Reverse-RESYNC

(1) Reverse-RESYNC Function/Quick Restore Function
The Reverse-RESYNC function is an extension of the RESYNC function of the ShadowImage. The Quick Restore function is a similar function with Reverse-RESYNC, but it is speeds up the operation.

When a pair in the Split status is requested to perform the Reverse-RESYNC, the differential data between the target volume and the source volume is copied to the source volume from the target volume.
When a pair in the Split status is requested to perform the Quick Restore, a volume map in DKC is changed to swap contents of Source volume and Target volume without copying the Source volume data to the Target volume. The Source volume and the Target volume are resynchronized when update copy operations are performed for pairs in the Duplex status.

Note on RAID Level and DCR swap:
The Quick Restore operation changes locations of the data for primary volumes and secondary volumes and location of DCR of ShadowImage pairs. Therefore, the operation may change RAID levels and HDD types of the volumes. For example, if the primary volume is RAID1 and the secondary volume is RAID5, Quick Restore operation changes the primary volume to RAID5 and the secondary to RAID1. RAID6 volume is also similar.
If you want to go back to the previous state, follow the actions below:
step1 : Stop HOST I/O to the pair
step2 : Split the pair
step3 : Perform Quick Restore for the pair
step4 : Restart HOST I/O to the pair

Due to the replacement of DCR setting locations, you must operate 1 or 2 shown below.
1. Set the same DCR location for Source volume and Target volume.
2. Reset the DCR settings of Source volume and Target volume before Quick Restore, and set DCR of Source volume and Target volume after the pair transits to the Duplex status by Quick Restore.

Unless you perform the operation above, I/O performance to the same data may be down for the change of the locations of cache-resident area after Quick Restore.

 Copyright © 2012, Hitachi, Ltd.

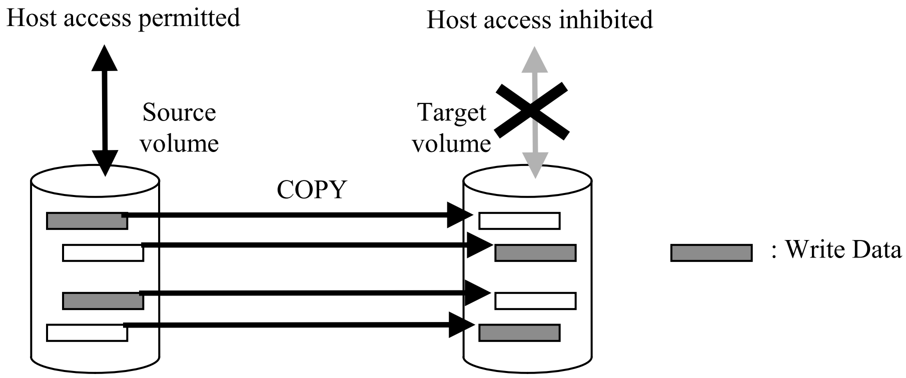


Fig. 3.11.6-1 Normal RESYNC Process

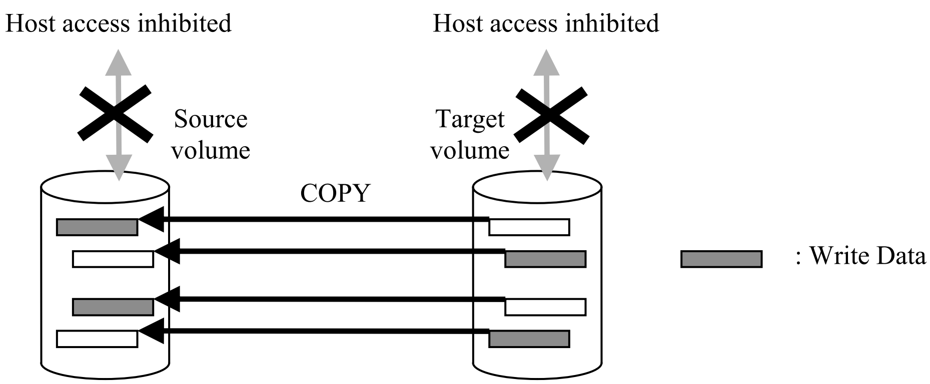


Fig. 3.11.6-2 Reverse RESYNC Process

 Copyright © 2012, Hitachi, Ltd.

(2) Specifications

No.	Item	Description
1	RESYNC copy pattern	• The data of the target volume is copied to the source volume. • The copy pattern can be selected by specifying a unit of operation. Specified operation unit : SVP/RMC: In units of pair operation at a time RAID manager: In units of command
2	Copy range	• In the case of the Reverse-Copy and Quick Restore in the Split status, a range for merging the writing into the source and target volumes
3	Copy format	• Same format as that of a copy in the Pair status
4	Applicable LDEV type	• SI : Normal volumes and CVSs and LUSE of them
5	Host access during copying	Source volume: Writing disabled Target volume: Reading and writing disabled NOTE: The reason why the source volume is not disabled to read is to make the volume recognizable by the host and it does not mean that the data is assured.
6	Specification method	• SVP/Storage Navigator: Add a specification for the RESYNC pattern onto the Pair Resync screen.
7	Conditions of command reception	• The pair concerned is in the Split status. • Another pair sharing the source volume is in the Suspend or Split status. → If this condition is not satisfied, the CMD RJT takes place. • When the Reverse-Resync or Quick Restore is being executed by another pair which is sharing the source volume, it is impossible to change the pair status of the pair concerned. (However, the pair deletion and pair suspension requests are excluded.) • The source volume of the pair concerned has no pair of the TC or in the Suspend status. (See Item No.14 in this table.)
8	Status display during copying	• SVP/Storage Navigator SI : COPY (RS-R) The display of the attribute, source or target, is not changed. • RAID manager Pair status display: RCPY The display of the attribute, source or target, is not changed.
9	Condition after normal end of the copy operation	• The pair concerned enters the Duplex status. • The conditions of the host access after the status transition are shown below. Source volume: Reading and writing enabled Target volume: Writing disabled

(To be continued)

Copyright © 2012, Hitachi, Ltd.

(Continued from the preceding page)

No.	Item	Description
10	Impacts on another pair	In another pair sharing the source volume, the part actually copied becomes the difference after executing this function. Example: Pair of the other target volumes in the 1:3 configuration
11	Operation when the copying terminates abnormally	(1) The pair concerned enters the Suspend status. (2) The source volume of the pair concerned is enabled to read and write. → Data is not assured. The target volume of the pair concerned is disabled to write. (3) The status of a pair sharing the source volume is not changed.
12	Operation when a suspension request is received during copying	Same as above
13	Relation to the cascade function	• The Reverse-RESYNC and Quick Restore cannot be executed for the L2 pair.

(To be continued)

Copyright © 2012, Hitachi, Ltd.

(Continued from the preceding page)

No.	Item	Description
14	Relation to the TrueCopy	• In the case where “M-volume of the TC” = “Source volume of the ShadowImage” “Pair status of the TC” = “Suspend” → The Reserve-Resync and Quick Restore can be executed. “Pair status of the TC” ≠ “Suspend” → The Reserve-Resync and Quick Restore cannot be executed. (Command Reject) • In the case where “R-volume of the TC” = “Source volume of the ShadowImage” “Pair status of the TC” = “Suspend” → The Reserve-Resync and Quick Restore can be executed. “Pair status of the TC” ≠ “Suspend” → The Reserve-Resync and Quick Restore cannot be executed. (Command Reject) • In the case where “Target volume of the ShadowImage” = “M-VOL of the TC” “Pair status of the TC” = “Suspend” → The Reserve-Resync and Quick Restore can be executed. “Pair status of the TC” ≠ “Suspend” → The Reserve-Resync and Quick Restore cannot be executed. (Command Reject) • A pair of the TC cannot be created with the volume of the ShadowImage executing the Reserve-Resync or Quick Restore. (Command Reject) These specifications do not depend on using external Volumes.

(To be continued)

 Copyright © 2012, Hitachi, Ltd.

(Continued from the preceding page)

No.	Item	Description
15	Relation to the Universal Replicator	• In the case where "Primary volume of the Universal Replicator" = "Source volume of the ShadowImage" "Pair status of the Universal Replicator" = "Suspend" → The Reserve-Resync and Quick Restore can be executed. "Pair status of the Universal Replicator" ≠ "Suspend" → The Reserve-Resync and Quick Restore cannot be executed. (Command Reject) • In the case where "Secondary volume of the Universal Replicator" = "Source volume of the ShadowImage" "Pair status of the Universal Replicator" = "Suspend" → The Reserve-Resync and Quick Restore can be executed. "Pair status of the Universal Replicator" ≠ "Suspend" → The Reserve-Resync and Quick Restore cannot be executed. (Command Reject) These specifications do not depend on using external Volumes.

Copyright © 2012, Hitachi, Ltd.

(3) Action to be taken when the pair is suspended during the Reverse-RESYNC
The recovery procedure to be used when the pair executing the Reverse-RESYNC is suspended owing to some problem or is explicitly transferred to the Suspend status by a command from the SVP/Web Console/RAID manager is explained below.

(a) Case 1: A case where the Suspend status can be recovered without recovering the LDEV concerned
This is equivalent to a case where the pair encounters an event that copying cannot be continued owing to a detection of pinned data or a staging time- out.
Or, it is equivalent to a case where the pair is explicitly transferred to the Suspend status by a command.
<<Recovery procedure>>

START

↓

Step 1: Delete the pair.

↓

Step 2: Remove the factor causing the suspension of the copying by referring to the SSB, etc..

↓

Step 3: Create a pair in the reverse direction (remove the source and target volumes).

↓

Step 4: Place the pair concerned in the Split status, and then delete the pair.

↓

Step 5: Create a pair again in the original direction.

↓

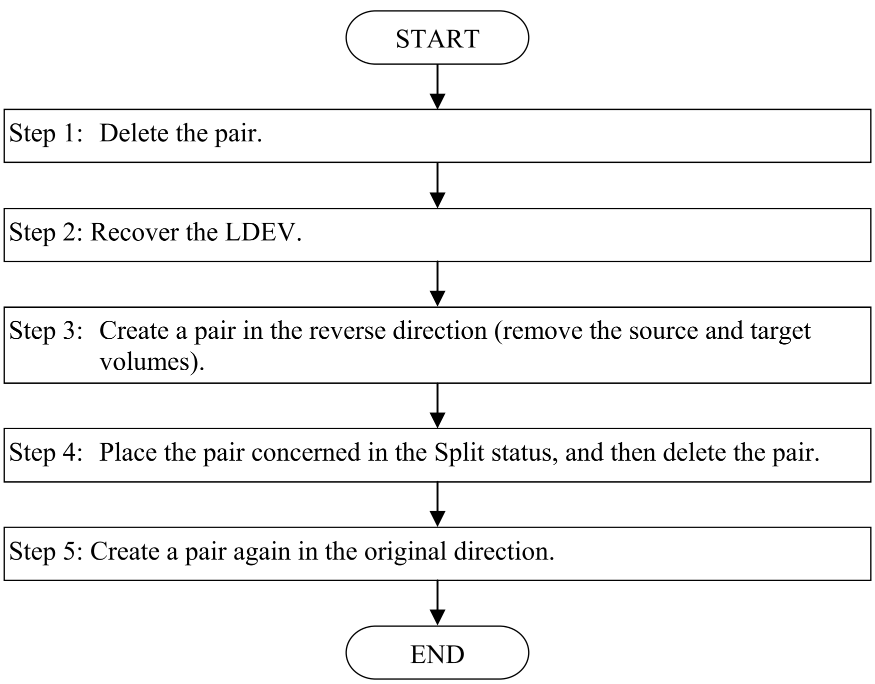
END

Copyright © 2012, Hitachi, Ltd.

(b) Case 2: A case where the Suspend status cannot be recovered unless the LDEV concerned is recovered
This is equivalent to a case that the LDEV is blocked.
To recover the blockade of the LDEV, an LDEV formatting or LDEV recovery is required. Both of them cannot be executed in the state that the ShadowImage pair is created. (A guard works against it.) Therefore, delete the pair once, recover the LDEV, and then create the pair once again.
However, in the pending state, caution must be taken because the data of the source volume is copied to the target volume if the pair is simply created again. Recover the blockade following the procedure below.

The following procedure is applicable just to a restoration of the source volume using the target volume. The following procedure does not include a procedure for directly restoring the source volume when the target volume is blocked.

• Case 2-1: A case where the source volume is blocked
<<Recovery procedure>>



Copyright © 2012, Hitachi, Ltd.

• Case 2-2: A case where the target volume is blocked

A recovery procedure for restoring data of the target volume is added because the copy source of the Reverse-RESYNC cannot be accessed.

<<Recovery procedure>>

START

↓

Step 1: Delete the pair.

↓

Step 2: Recover the LDEV.

↓

Step 3: Create the data of the target volume again (restore it again or anything).

↓

Step 4: Create a pair in the reverse direction (reverse the source and target volumes).

↓

Step 5: Place the pair concerned in the Split status, and then delete the pair.

↓

Step 6: Create a pair again in the original direction.

↓

END

 Copyright © 2012, 2013, Hitachi, Ltd.

3.11.7 Notes on powering off

When performing a powering off, take notice of the following.

Item	Note	Reason

1	(ShadowImage)
	Take care that the time required for the copying becomes longer. Make a schedule taking the copying time into consideration.
		If data in the shared memory has volatilized when the next powering on is performed, the following phenomena occur.
		• When the pair is in the Pending or Resync status, the data, from which a copying has been completed before the powering off, is also treated as data to be copied again.
		  • Even if no I/O has been issued, the rate of data identity does not reach 100% when the pair status is changed to Duplex.
		  • The data that has become the one to be copied again is copied to the secondary volume after the pair status is changed to Duplex.
		• When the pair is in the Duplex status, the data, from which a copying has been completed before the powering off, is also treated as data to be copied again.
		  • The rate of data identity will be 0%.
		  • The copying of the data, which has become the one to be copied again, is performed in the state in which the pair is in the Duplex status.
		• When the pair is in the Split status, the whole volume will be a differential between the two volumes.
		  • The rate of data identity will be 0%.
		  • Data of the whole volume is copied when a resynchronization is performed.

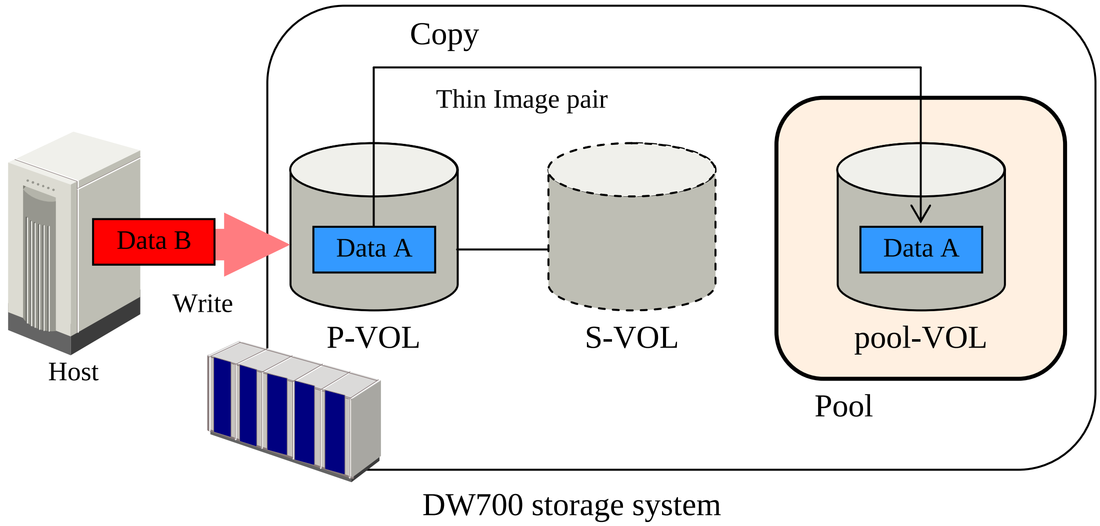
2	(ShadowImage) As to a pair in the Split transitional status (SP-Pend, V-Split), complete the copying of it and put it in the Split status.	If data in the shared memory has volatilized when the next powering on is performed, the following phenomenon occurs. • When the pair is Split transitional status (SP-Pend, V-Split), it is changed to Suspend.
3	(ShadowImage At-Time Split function) Perform PS OFF when by split operation by At-Time Split function, after change status of all pairs belonging to consistency group is completed.	Pair by which change status is not carried out even if it carries out PS ON may occur.

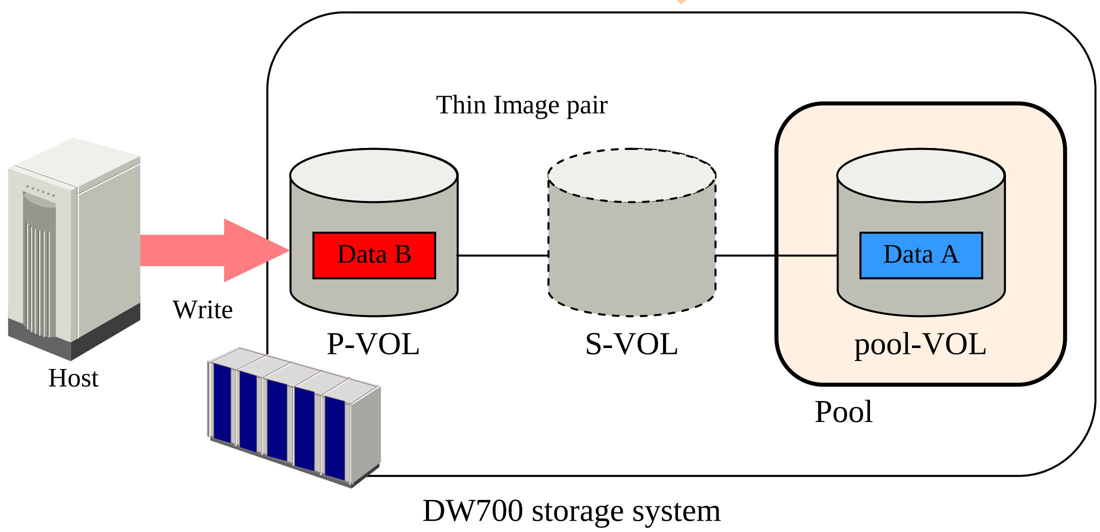
Copyright © 2012, Hitachi, Ltd.

3.11.8 Thin Image option

Thin Image option requires the shared memory. For more details, refer to INSTALLATION SECTION "6.1 Required CM Capacity" (INST06-10).

(1) Thin Image function
It is a product to copy and manage the data in the storage system as well as the ShadowImage function. The Thin Image forms the pair of which the logical volume is made as the primary volume (hereafter, indicated as P-VOL), and the virtual volume (hereafter, indicated as V-VOL) is made as the secondary volume (hereafter, indicated as S-VOL).
Because the S-VOL of the Thin Image pair is the V-VOL with no reality, the S-VOL does not actually consume the capacity of the storage system.
As for the Thin Image pair, only the part to be updated is copied to the pool-VOL among the data of the P-VOL. Therefore, the capacity used by the entire storage system can be reduced.
It is possible to access the S-VOL of the Thin Image pair. In this case, the data of the P-VOL is referred to via the S-VOL. Therefore, the load concentrates on the parity group of the P-VOL. Also, if the P-VOL becomes a failure, it cannot access the S-VOL.





(2) Pool-VOL

A pool is an area to store the snapshot data acquired by the Thin Image.
The pool is configured with two or more pool-VOLs, and the snapshot data is actually stored in the pool-VOL.
As a result of writing the data in the volume of the Thin Image pair, when the amount of the pool in use exceeds the capacity of the pool, the Thin Image pair becomes PSUE (status of the failure occurrence), and cannot acquire the snapshot data.

Copyright © 2012, Hitachi, Ltd.

3.12 (Blank)

Copyright © 2012, Hitachi, Ltd.

3.13 Volume Migration

3.13.1 Volume Migration Overview

This document describes the function of Volume Migration that is one of program products.

RAID system can be constructed by several types of physical drives and three types of RAID levels (RAID1, RAID5 and RAID6).

This combination of the type of physical drive and the type of RAID level provides a system that cost and performance are optimized to user environment. However, it is difficult to get information about actual operation of physical drives in the RAID system unlike other storage systems.

(1) Volume Migration provides solutions of the problem and supports decision of users to determine system construction as described below.
  (a) Load balancing of system resources
    Unbalance of utilization of system resources makes performance worse. Volume Migration supports decision of optimized allocation of logical volumes to physical drives.
  (b) Migration of logical volumes optimized to access patterns to physical drives
    For instance, RAID5/RAID6 are suitable to sequential access, and RAID1 of high performance drive is suitable to random access that is required small response time. Volume Migration shows types of access pattern to physical drives clearly, and supports migration of logical volumes to suit the access pattern.

(2) Volume Migration consists of following subfunction to achieve above purposes. Users can refer to utilization of system resources monitored by monitor function, decide reallocation plan by using estimate function, and reallocate the logical volumes by volume moving (migration) function.
  (a) Monitor function(*1) monitors and shows utilization of system resources.
  (b) Estimate function estimates utilization of parity groups after migration of logical volumes.
  (c) Volume moving (migration) function moves logical volumes to specified parity groups.
  (d) Automating Migration function makes a migration plan from information that users preset, and moves the logical volumes by the migration plan automatically.

*1: The license of Performance Monitor is required.
For details, refer to "Performance Guide" of Program products.

Copyright © 2012, Hitachi, Ltd.

3.13.2 Hardware requirements

Addition of SM capacity setting may be needed if SM capacity setting in the DKC is not enough. Please refer to INSTALLATION SECTION.

3.13.3 Software requirements

The Program Product, Performance Monitor is required.

 Copyright © 2012, Hitachi, Ltd.

3.13.4 Monitor function

The Performance Monitor Program is required to perform this function.

(1) How to start and stop.
Monitoring starts in DKC by direction from the Monitoring Option window.

Monitor function monitors ratios of utilization of system resources described below.
(a) Usage rates of channel processors
(b) Starnet utilization ratio
(c) Usage rates of DDRs
(d) Parity group utilization: Disk utilization of parity groups.
The used time of physical drives in a parity groups.
(e) Parity group utilization ratio of each logical volume: The used time of physical drives of synchronous and asynchronous access on each logical volume, averaged by the number of physical drives in the parity group.
Parity group utilization means the sum of utilization of each logical volume in the parity group.

Directions for an stop from the Monitoring Option window to finish the collection of the Monitoring information.

(2) How to collect
Monitoring information is acquired automatically until directions for an acquisition end are issued after directions for an acquisition start of the Monitoring information are issued on the Monitoring Option Window.
When acquired Monitoring information is collected in SVP, it learns to refer to it with a [Physical] tab.
It is done automatically collecting opportunity Monitoring information every 15 minutes.
The collected Monitoring information up to 3 months is stored in the hard disk of SVP.
When a monitor function is continued for more than 3 months, it is erased in order from the old information, and new Monitoring information is cumulated.

(3) How to view
The Monitoring information which accumulated in SVP can be referred to with a [Physical] tab.
The period of the Monitoring information which accumulated can be specified by the operation of the part Term.(It is optional in the unit for 15 minutes.)
Information about the Monitoring information of the period specified in the part Term is indicated in the Table part and the Graph part.
The utilization of the resource (average and maximum value) chosen in the Tree part is indicated in the Table part.
The utilization of the element chosen by the Table part is graphed and indicated by the Graph part.

Copyright © 2012, Hitachi, Ltd.

3.13.5 Estimate function

The estimate function estimates changes of parity group utilization and parity group utilization of each logical volume after migration of the logical volume to specified parity group. The estimate function estimates the changes from the monitored information.

Copyright © 2012, Hitachi, Ltd.

3.13.6 Volume moving (migration) function

Volume moving (migration) function moves data in logical volume (source volume) to physical location of another logical volume (destination volume). Users specify the volumes in Volume Migration utility window.

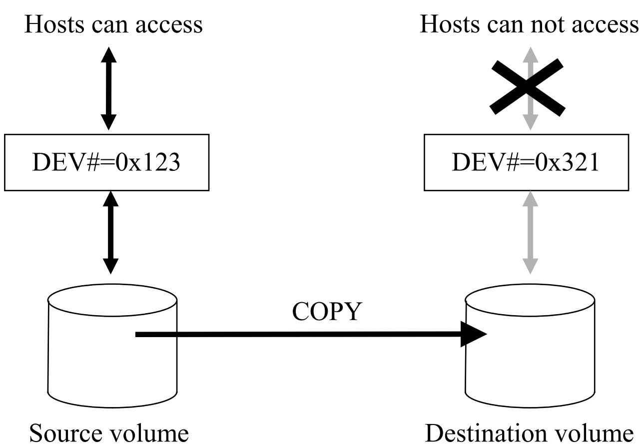


Fig. 3.13.6-1 Before moving

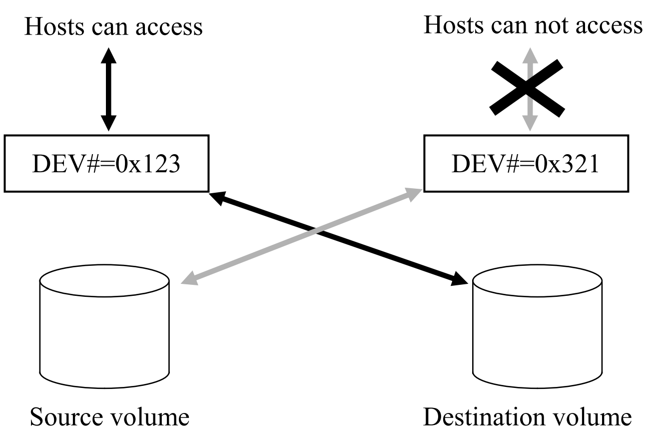


Fig. 3.13.6-2 After moving

Copyright © 2012, Hitachi, Ltd.

(1) Volume moving function Overview

(a) Source volumes:

The following volumes cannot be used as source volumes:

- Volumes which are set as command devices (devices reserved for use by the host)
- Volumes which have FlashAccess (also called DCR) data stored in cache
- Volumes which are in an abnormal or inaccessible condition (e.g., pinned track, fenced)
- Journal Volumes which are used by Universal Replicator
- Volumes which are reserved by a migration program other than Volume Migration
- Volumes on which CCI is performing volume migration operation
- Volumes that form a Thin Image pair
- Virtual volumes and pool volumes
- Volume executing Quick Format

If the status of volumes that forms TrueCopy pairs is PSUS or PSUE or PAIR, the volumes can be used as source volumes. If not, the volumes cannot be used as source volumes. If you delete an TrueCopy pair from an MCU, the status of the P-VOL and the S-VOL changes to SMPL so that the volumes can be used as source volumes. If you delete an TrueCopy pair from an RCU, the status of the P-VOL changes to PSUS and the status of the S-VOL changes to SMPL, so that the volumes can be used as source volumes.

— Dynamic Provisioning Volume

If the Dynamic Provisioning Volume are used as source volumes, the Dynamic Provisioning Volume that associates same pool volume as source volumes can’t be specified as target volumes.

The Dynamic Provisioning V-VOL whose capacity is being expanded cannot be used as the source volume for volume migration. After DP V-VOL expansion is completed, the V-VOL can be used as the source volume for migration. In this case, make sure that the capacity of the target volume must be the same as that of the expanded source volume.

 Copyright © 2012, Hitachi, Ltd.

— Volumes which configure the Universal Replicator pairs
If the status of the volumes is COPY or PAIR, they cannot be source volumes.
Also, if the volumes which configure the Universal Replicator pairs are used as source volumes, external volumes cannot be specified as target volumes.

When Delta Resync is set in 3D Multi Target Configuration (See Fig. 3.13.6-3), P-VOL and S-VOL of a pair of Universal Replicator and for Delta Resync can be set as a source volume. However, the status of each pair in 3D Multi Target Configuration is required to be as the following table.

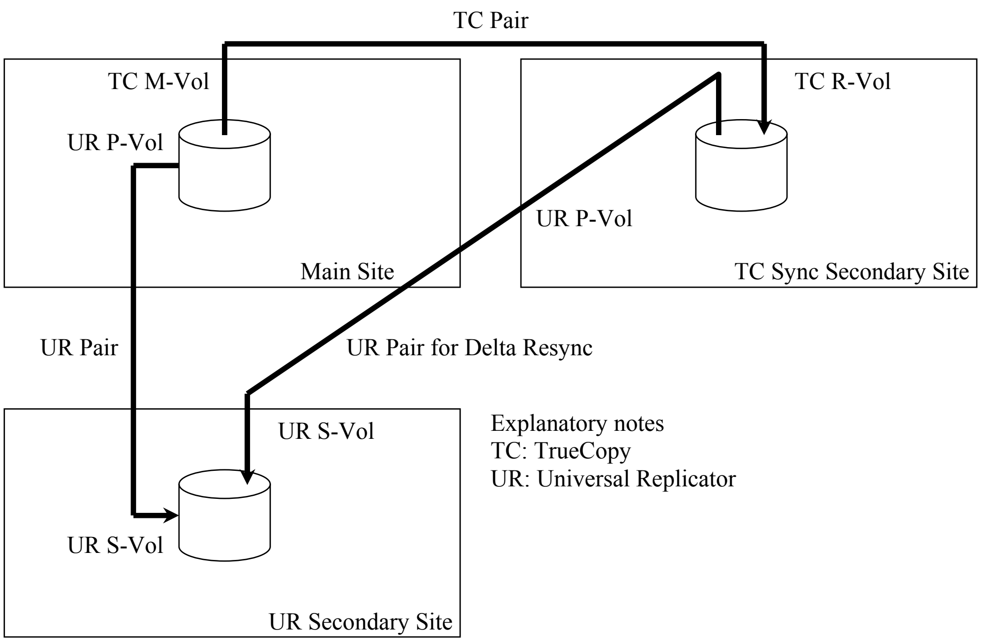


Fig. 3.13.6-3 3D Multi Target Configuration (Universal Replicator)

Table 3.13.6-1 The Status of Each Pair when P-VOL of UR Pair for Delta Resync is Set as a Source Volume

Pair	Pair Status
TC Pair	PSUS
UR Pair	Random Pair Status
UR Pair for Delta Resync	HOLD or HLDE

Table 3.13.6-2 The Status of Each Pair when S-VOL of UR Pair for Delta Resync is Set as a Source Volume

Pair	Pair Status
TC Pair	Random Pair Status
UR Pair	PSUS
UR Pair for Delta Resync	HOLD or HLDE

Copyright © 2012, Hitachi, Ltd.

For volumes that form an ShadowImage pair, it depends on the status or configuration of the pair whether the volumes can be used as source volumes, as explained below:

- If the status of the pair is not SP-Pend, the volumes can be used as source volumes. If the status of the pair is SP-Pend, the volumes cannot be used as source volumes.
- The table below explains whether volumes that do not form a cascade pair can be used as source volumes:

Table 3.13.6-3 Whether volumes that do not form a cascade pair can be used as source volumes

If the pair is configured as follows	Can P-VOLs be used as source volumes?	Can S-VOLs be used as source volumes?
If the ratio of P-VOLs to S-VOLs is 1:1	Yes	Yes
If the ratio of P-VOLs to S-VOLs is 1:2	Yes	Yes
If the ratio of P-VOLs to S-VOLs is 1:3	No	Yes

- The table below explains whether volumes that form a cascade pair can be used as source volumes:

Copyright © 2012, Hitachi, Ltd.

Table 3.13.6-4 Whether volumes that form a cascade pair can be used as source volumes

If the pair is configured as follows	Can P-VOLs be used as source volumes?	Can S-VOLs be used as source volumes?
If the pair is an L1 pair and the ratio of P-VOLs to S-VOLs is 1:1	Yes	Yes
If the pair is an L1 pair and the ratio of P-VOLs to S-VOLs is 1:2	Yes	Yes
If the pair is an L1 pair and the ratio of P-VOLs to S-VOLs is 1:3	No	Yes
If the pair is an L2 pair and the ratio of P-VOLs to S-VOLs is 1:1	Yes	No
If the pair is an L2 pair and the ratio of P-VOLs to S-VOLs is 1:2	No	No

NOTE: If any of the following operations is performed on a source volume, the volume migration process stops:

- TrueCopy operation that changes the volume status to something other than suspended
- ShadowImage operation that changes the volume status to SP-Pend.
- Universal Replicator operation that make the volumes the status of COPY

Copyright © 2012, Hitachi, Ltd.

(b) Target volumes:

Target volumes must be reserved prior to migration. The Volume Migration Web Console software allows you to reserve volumes as Volume Migration target volumes.
Hosts cannot access Volume Migration-reserved volumes.
The following volumes cannot be reserved as target volumes:

- Logical Unit Size Expansion (LUSE) volumes
- Volumes which are set as command devices (devices reserved for use by the host)
- Volumes which are assigned to ShadowImage or Remote Copy pairs
- Volumes which are reserved for ShadowImage operations
- Volumes which have FlashAccess (also called DCR) data stored in cache
- Volumes which are in an abnormal or inaccessible condition (e.g., pinned track, fenced)
- Onlined volume
- Volumes which are used by Universal Replicator (Primary Data Volume, Secondary Data Volume and Journal Volume)
- Volumes which are set as Read Only, Protect or S-VOL Disable from Data Retention Utility
- Volumes which are reserved by a migration program other than Volume Migration
- Volumes on which CCI is performing volume migration operation
- Volumes that form a Thin Image pair
- Virtual volumes and pool volumes
- Volume executing Quick Format
- The Dynamic Provisioning V-VOL whose capacity is being expanded

 Copyright © 2012, Hitachi, Ltd.

(c) Specifying Volumes:
Source volumes and Target volumes are specified by LDEV number.
If the volumes are set as LUSE, you can specify by one LDEV number. Volume Migration will migrate the LDEV which is one of a LUSE and its access is higher.

(d) Moving of multi volumes:
Moving of volumes can be performed by repeating instruction about each volume.

In using Volume Migration for manual volume migration, the number of migration plans that can be executed concurrently might be restricted. The number of migration plans that can be executed concurrently depends on the following conditions.

- How much shared memory is available for differential tables:
  You may use 26,176 differential tables if no additional shared memory for differential tables is installed.
  You may use 104,768, or 209,600 differential tables if additional shared memory for differential tables is installed.

- How much shared memory is available for pair tables:
  You may use 8,192 pair tables if no additional shared memory for differential tables is installed.
  You may use 16,384 pair tables if additional shared memory for differential tables is installed.
  To install additional shared memory for differential tables, please call the Support Center.

- Capacity of each volume to be migrated:
  The number of differential tables, pair tables needed to migrate one volume differs size of the volume. For the number of differential tables, pair tables needed for migrating an volume, see section (i).

 Copyright © 2012, Hitachi, Ltd.

You can estimate the maximum number of migration plans that can be executed concurrently by applying the above conditions into the following equation:

$\Sigma\ (\alpha) \leq (\beta)$

and

$\Sigma\ (\gamma) \leq (\delta)$

$\Sigma\ (\alpha)$ stands for the total number of differential tables needed for migrating all volumes (β) stands for the number of differential tables available in the DKC710I storage system.
$\Sigma\ (\gamma)$ stands for the total number of pair tables needed for migrating all volumes (δ) stands for the number of pair tables available in the DKC710I storage system.
For example, if you want to create 20 migration plans of size capacity of the volumes (size of a volume is 3,019,898,880 KB), the number of the required differential tables is 577, the number of the required pair tables is 17 that can be found by the calculation described in section (i). When you apply this number to the equation, it will be as follows:
$[\ (577 \times 20) = 11{,}570\ ] \leq 26{,}176$
and
$[\ (17 \times 20) = 350] \leq 8{,}192$
Since this equation is true, you can create all the migration plans that you wish to create.

In this section, we mentioned the calculation of the maximum number of migration plans when only Volume Migration is running. However, in fact, the total number of differential tables used by ShadowImage and Volume Migration should be within the value of (β), and the total number of pair tables used by ShadowImage, Thin Image, and Volume Migration should be within the value of (δ). For details on how to calculate the number of differential tables, pair tables used by the programs other than Volume Migration, please refer to the following manuals:

- For ShadowImage, please refer to “Hitachi ShadowImage® User Guide”.
- For Thin Image, please refer to “Hitachi Thin Image User Guide”.

Copyright © 2012, Hitachi, Ltd.

(i) Calculating Differential Tables, Pair Tables Required for Volume Migration
When you migrate volumes, use the expression in following table to calculate the total number of the required differential tables, pair tables per migration plan.

Total number of the differential tables per migration plan = ((X) ÷ 256) ÷ (Z)
(X): The capacity of the volume to be migrated. (KB) (*1)
(Z): The number of the slots that can be managed by a differential table. (20,448)

Note that you should round up the number to the nearest whole number.
For example, when provided that the number of the cylinders is 3,019,898,880 kilobytes ((X) in the expression above), the calculation of the total number of the differential tables is as follows.
(3,019,898,880 ÷ 256) ÷ 20,448 = 576.9014085
When you round up 576.9014085 to the nearest whole number, it becomes 577.
Therefore, the total number of the differential tables for one migration plan is 577.
In addition, 1 pair tables per 36 differential tables is used. The number of pair tables used for above- mentioned 17.

*1: If the volume is divided by the VLL function, this value means the capacity of the divided volume.

Copyright © 2012, Hitachi, Ltd.

(e) Abort moving:
Users can direct to abort the instructed moving before completion.
With aborting, the data in the destination volume is not guaranteed.
Users can direct to abort by each LDEV.

(f) Notice when the DKC is maintenance:
Volume migration may be failed if you start the following action:
(i) Replace Cache/HDD
(ii) Install/deinstall Cache/HDD
(iii) Change of SM capacity setting

(2) Conditions for moving
Data moving is performed when all conditions about source volume and destination volume described bellow are satisfied.
(a) Both of the volumes have same size.
(b) There is no PIN data in the source volume.
(c) Both of the volumes are not blockade.
(d) Both of the volumes in same DKC.
(e) The volumes are not instructed to move already and not waiting to move.
(f) The volumes are not combination of CVS Volume and Normal Volume.

(3) Viewing History
Users can see the history of volume moving (migration).

Copyright © 2012, Hitachi, Ltd.

3.13.7 Decision of volume moving (migration)

Volume Migration supports decision of users about disk system performance tuning by logical volume moving (migration). This section describes usage and points to notice about monitor function.

(1) Inspecting utilization of system resources
First of all, using monitoring function, a user investigates MP utilization, Startnet utilization, DRR utilization, RAID group utilization. a user investigates whether there exists overloaded resources, or imbalance of resource utilization. Then the user tunes resource utilization in the manner described in the following clause.

NOTE: Due to average system resource utilization, there will be such a case as a portion of system performance will be negatively effected although total performance of a system will be improved. For example, if there exists RAID groups A and B of utilization 20% and 90% respectively, and if the utilization will become 55% and 55% if a logical volume residing in parity group B moves parity group A. Then response time of I/Os to parity group A will be increased while response time of I/Os and throughput to parity group B will be improved.

(2) Tuning Starnet utilization
Since Starnet are common resources in a storage system, migration of logical volumes does not improve system performance. The user should consider migration the data to other DKC.

(3) Tuning MP utilization
Migration of logical volumes does not improve MP performance. Therefore if MPs are overloaded on an average, the user should consider installation of new MPs. And if the utilization of MPs are imbalance, the user should consider that channel paths connecting to a CHB, which includes overloaded MPs, is reconfigured into the connection to another CHB which includes MPs of lower utilization.

(4) Tuning DRR utilization
If utilization of DRRs is in high average, the user should consider installing new DKBs and HDDs or migration logical volumes of RAID5/6 to RAID1 or migration the data to other DKC. After installation of DKBs and HDDs, logical volumes which had high traffic of write access (especially of sequential write access) should be migrated to a parity group in newly installed HDDs. The estimate function cannot simulate the DRR utilization.

(5) Tuning RAID group utilization
If parity groups are in high utilization, the user should consider installing new HDDs. After installation of HDDs, logical volume which had high traffic of I/Os should be migrated to a parity group in newly installed HDDs by reference RAID group utilization on a logical volume.

If utilization of each parity group is imbalanced, the user should consider migration of logical volumes from the current parity group showing high utilization to the one showing lower utilization.

These methods should be applied with a view to large improvement. There would be least improvement in a case of slight difference of utilization of each parity group, or if DRRs or DKPs are already comparatively highly utilized.

If a number of the condition are right, the user should decide to do considering each examination items.

When errors exist in the system, utilization of system resource can increase or be unbalanced.

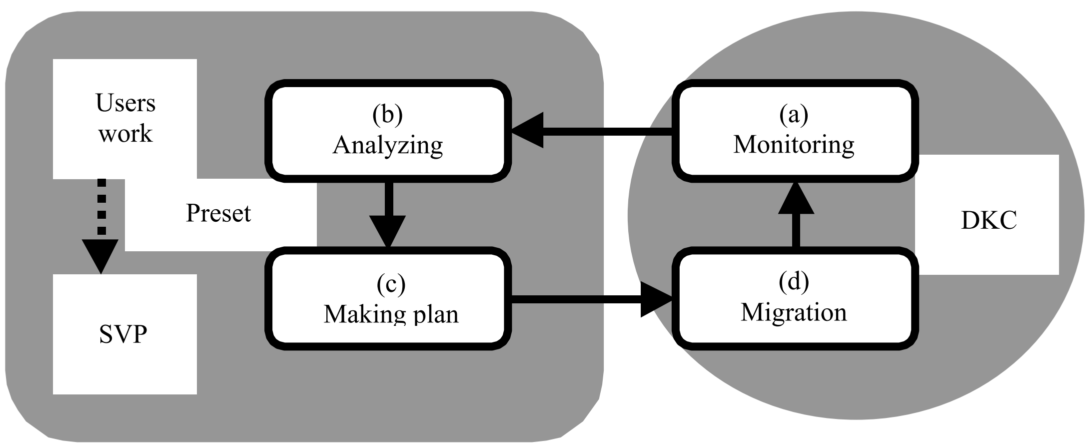
Copyright © 2012, Hitachi, Ltd.

3.13.8 Automating Volume Migration function

(1) Automating Volume Migration function Overview
Automating Volume Migration function provides a typical tuning method of Volume Migration based on parameters given by users, and performs tuning plan automatically.

(2) Overview of tuning
Tuning by Volume Migration can be proceeded by performing the following steps repeatedly.
(a) Monitoring information
(b) Analyzing information
(c) Making volume migration plan (decision of volume migration)
(d) Moving volume (migration)
(a’) Monitoring information again to confirm condition and effect of the performed tuning.

Preset function reduces users’ work by providing a typical tuning method of Volume Migration based on parameters given by users, and performs tuning plan automatically. When (b) Analyzing and (c) Making plan are done by users, it can perform fine tuning.



Copyright © 2012, Hitachi, Ltd.

(3) Process of tuning by Automating Volume Migration function

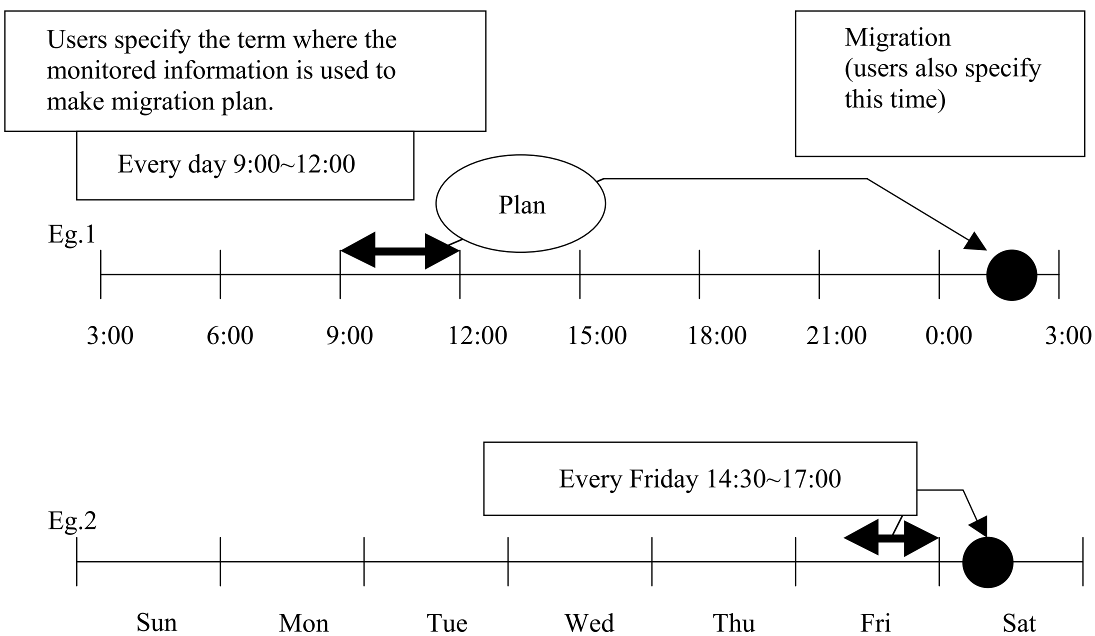
(a) Basic process flow of tuning by Automating Volume Migration function

Preset function performs the following two processes.

- Making Plan: Detecting volumes that have problems by monitored information (disk utilization) and making plan of volume Migration
- Migration: Moving volume according to the plan.

These two processes are repeated by the cycle described above, the length of the cycle depends on users. In the preset function, this cycle is supposed from one day to one week or several weeks at most.

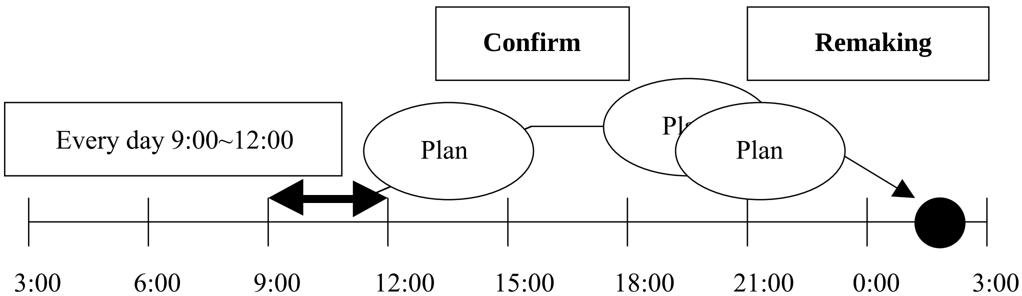
Users can specify the term where the monitored information is used to make the plan focusing on concerns of users. For example, users can specify the highest load term in a day or in a week as the referred term for making plan.



Copyright © 2012, Hitachi, Ltd.

(b) Confirming and remaking of plan by users
During the period from the planning by preset function until migration, users can see the plan and delete it if needed.

During the same period as above, users can remake the plan manually with changing parameters.



Copyright © 2012, Hitachi, Ltd.

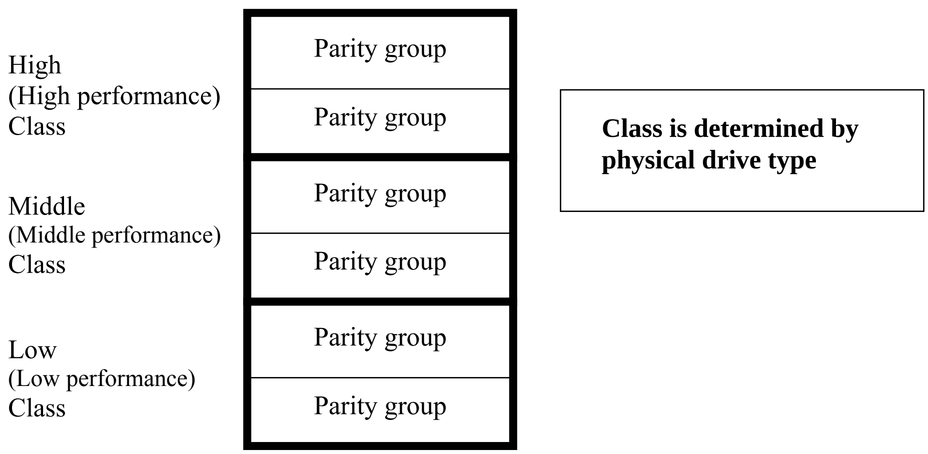
(4) Making plan

(a) Tuning based on disk utilization ratio

Volume Migration monitors utilization ratios of various system resources, but the preset function uses disk utilization ratio to make a migration plan to solve disk neck as a typical tuning method. Users could refer to other information if needed.
Users also specify parameters for tuning based on disk utilization ratio. Users should set these parameters for their system (preset function provides default value for these parameters roughly).

(b) Hierarchy of parity groups and management by class

Parity groups in DKC have hierarchy by drive type and RAID type. Volume Migration provides a function to optimize the usage of this hierarchy. Preset function manages this hierarchy as class (parity group set).
Parity groups are divided into classes. The classes are ordered from high level (high performance) to low level (low performance). This classification is decided by performance of physical drive type of each parity group.



A method of migration plan which the preset function performs is described below. Preset function uses the monitored information in the term that is the data to be referred to make the plan.

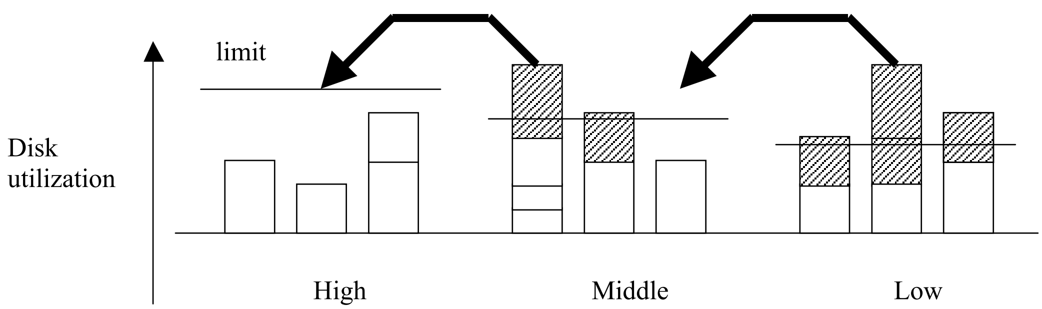
Copyright © 2012, Hitachi, Ltd.

(b-1) Management by maximum limit of disk utilization ratio

A maximum limit of disk utilization (parity group utilization) is specified to each parity group. Users should specify this limit for their systems. (Volume Migration uses the default value but it is desirable that user sets the desired value.)

For the parity groups that exceed this limit, preset function makes plan of moving volumes from this parity group to another parity group in higher class.

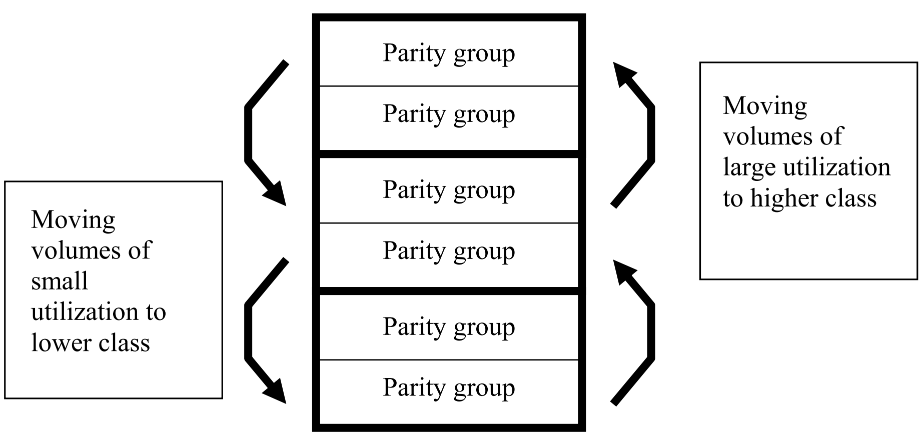
This avoids physical disk neck and provides load balancing of disk utilization.



(b-2) Selection of volumes to migration

In the migration plan described above, volumes of larger disk utilization are selected to be moved from the parity group that exceeds the limit to higher class. Moving larger utilization volumes to higher performance class is expected to makes large tuning effect. And larger utilization means larger amount of access from host, and this also makes large tuning effect.

When the reserved (empty) volumes run short in the high class, volumes of smaller disk utilization are selected to be moved from higher class to lower class to make reserve volumes.



Preset function provides some criteria for you to select them, average of disk utilization, average of highest Nth value of disk utilization in the referred term, and a value considering sequential/random access pattern.

(b-3) Specifying maximum limit of disk utilization ratio by users

When the same maximum limit of disk utilization ratio described above is given to each class (parity group) classified by drive types, performance of each physical drive type makes performance of each class directly.
When the users specify the limit to each class with bias, users can make the difference of performance of classes larger or smaller.
Users can specify parity groups to fixed parity groups in which volumes users do not want to move automatically. Preset function does not make a migration plan about fixed parity groups and volumes in the fixed parity groups.

(b-4) Notice for making plan

Volume Migration can make these plans only on the following conditions:

- Volume Migration can estimate the disk utilization ratio against all migrated parity groups.
- The disk utilization rate of all migrated parity groups should not be over the maximum rate. If the rate of one parity group is over the maximum rate, Volume Migration could not make a plan.

(b-5) Notice for reference term

Volume Migration could not use old information before the last volume migration in order to reduce the influence of performance by volume migration.
Therefore, Volume Migration sometimes fails in making a plan by lack of information.

(5) Moving (migration) by preset function

Preset function performs moving (migration) process once a day at time specified by users.
If there is a migration plan made by preset function at the time, volumes are moved by the plan.
Users can specify the limit of moving to avoid the overload by moving (data copy) process.
If the disk utilization ratio of parity groups in which moving is started exceeds the moving process limit, the moving in the parity group will be aborted.
Users can specify the time limit of moving. If a plan does not complete in the time limit, the remaining plan will be executed in the following day. If the new plan will be made until the next migration time, those remaining plans will be deleted before making the plan.

 Copyright © 2012, Hitachi, Ltd.

3.14 Data Assurance at a Time When a Power Failure Occurs

Refer to “2.2.3 (5) Battery” (THEORY02-110) for the operation to be done when a power failure occurs.

Copyright © 2012, Hitachi, Ltd.

3.15 Universal Volume Manager (UVM)

3.15.1 Overview

Universal Volume Manager is a program product that realizes the virtualization of the storage system. Universal Volume Manager enables you to operate multiple storage systems including DW700 storage system as if they are all in one storage system. As Universal Volume Manager realizes the virtualization of the storage system, the system administrator can manage the different kinds of multiple storage systems as one storage system.
Once you connect the DW700 storage system and another kind of storage system using Universal Volume Manager, the system administrator can also manage another storage system using DW700 Storage Navigator. For example, the system administrator can set the path from a host computer to the volume of the external storage system using LUN Management of DW700.
In addition to the function of Universal Volume Manager, the ShadowImage function enables you to easily make a backup copy of data stored in the DW700 storage system to another storage system. It is also easy to restore the backed up copy of data to DW700 storage system.
In this manual, the source DW700 storage system is called “local storage system”, and the connected storage system is called “external storage system”. The volume managed in the local storage system is called “internal volume”, and the volume in the external storage system is called “external volume”.

Features of Universal Volume Manager are as follows:
By mapping an external volume as an internal volume using Universal Volume Manager, it becomes possible to operate the external volume using DW700 Storage Navigator as if it is a volume in the DW700 storage system.
“Mapping” means assigning the CU:LDEV numbers of the internal volumes to the external volumes. By assigning the numbers of the internal volumes to the external volumes, the system administrator will be able to operate not only internal volumes of DW700 storage system but also external volumes using DW700 Storage Navigator.

Copyright © 2012, Hitachi, Ltd.

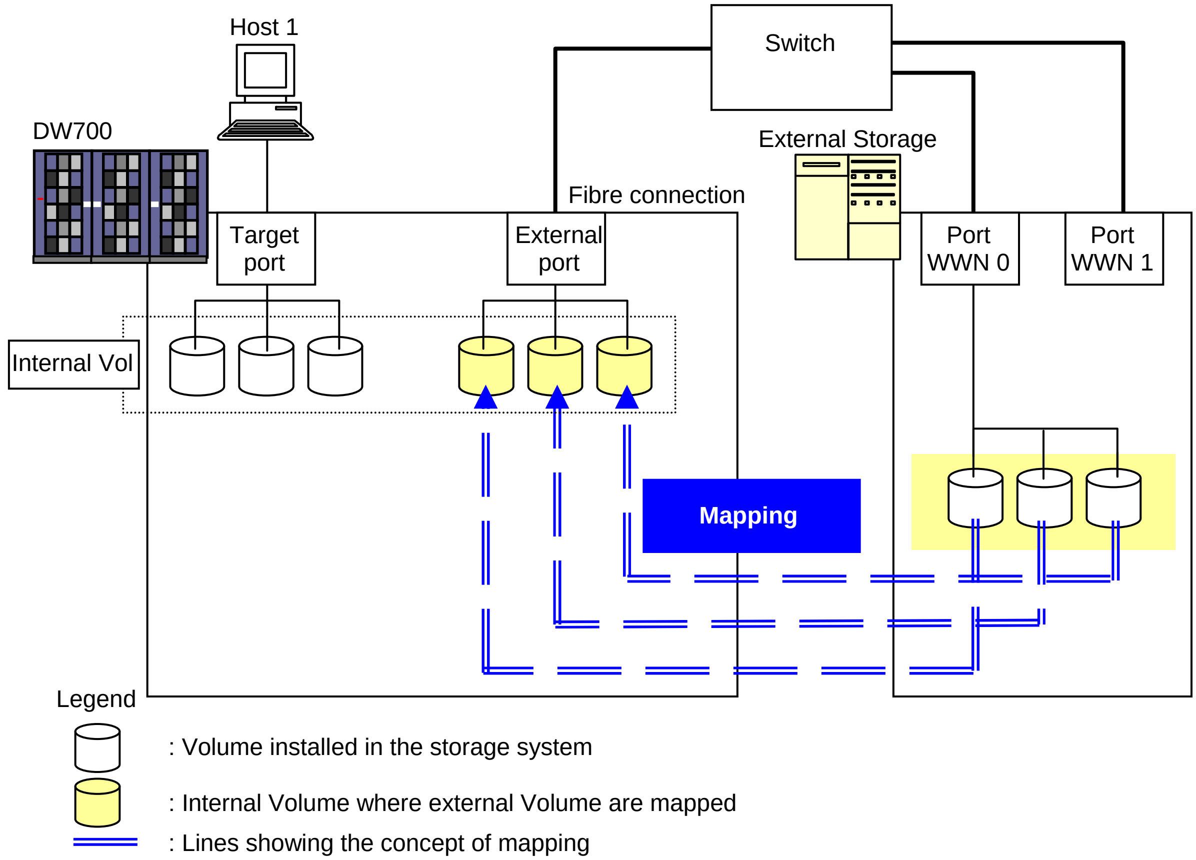


Figure 3.15.1-1

Figure 3.15.1-1 shows the idea of connection between a DW700 storage system and an external storage system which are connected by the Universal Volume Manager function. In Figure 3.15.1-1, the external storage system is connected to the external port of the DW700 storage system via a switch using the Fibre-channel interface. The external port is a kind of port attribute, which is used for Universal Volume Manager. In Figure 3.15.1-1, the external volumes are mapped as DW700 volumes.

Copyright © 2012, Hitachi, Ltd.

3.15.2 Procedure of using external volumes

(1) Prepare in the external storage system a volume to be used for UVM.
(2) Change the port attribute to External.
(3) Connect the external storage system to the external port.
(4) Search for the external storage system from the UVM operation panel (Discovery).
(5) Map an external volume.
  (a) Register to an external volume group
  (b) Select the cache mode
(6) Define the host path.
(7) Other settings.

3.15.2.1 Prepare in the external storage system a volume to be used for UVM

Prepare a volume to be used for UVM in the external storage system connected to the DW700.

NOTE: The volume in the external storage system should be about 47 MB (96,000 Blocks) ~ 60 TB (128,849,018,880 Blocks). If the capacity of the external volume is 60 TB or larger, you can use up to 60 TB of the volume.
If one mapped external volume uses as one internal volume, the external volume size must be less than 128,849,011,200 blocks or smaller.

Copyright © 2012, Hitachi, Ltd.

3.15.2.2 Change the port attribute to External

The port used for Universal Volume Manager needs to be set as the external port. When the external storage system is connected to the external port of the local storage system, you can view the information on the external storage system from the DW700 Storage Navigator computer. The external storage system cannot be connected to the ports other than the external port.
In order to set the port attribute to external, you need to release the paths set to the port. The attribute of the port where the paths are set cannot be changed to external. Before starting the Universal Volume Manager operations, you need to know the ports whose attributes can be changed to external.

NOTE: The ports whose attributes are set for remote copy software (eg., RCU target, initiator) or the other features cannot be used as external ports for Universal Volume Manager. Change the port attribute to external if the port attribute is set to other than external.

3.15.2.3 Connect the external storage system to the external port

Insert a Fibre cable into the external port from the external storage system.

3.15.2.4 Search for the external storage system from the UVM operation panel (Discovery)

Select [Create External Path Group] in ‘ADD External Volumes’ window. Then select [Discover External Target Ports] in ‘Create External Path Group’ window.
You cannot use the external volume from the DW700 by just discovering the external storage system. You need to perform the Add LU operation (mapping) shown in the next section.

Copyright © 2012, Hitachi, Ltd.

3.15.2.5 Map an external volume

When you connect the external storage system to the external port, volumes in the external storage system (external volumes) can be mapped as volumes in the local storage system (internal volumes). Confirm which volumes in which external storage system you need to map as internal volumes.
Only one external volume can be mapped as one internal volume. The maximum number of external volumes, which can be mapped, is 4,096 per port.
When the external volume is more than 60 TB (128,849,018,880 Blocks), you can access the data stored in the field up to 60 TB. If one mapped external volume uses as one internal volume, the external volume size must be less than 128,849,011,200 blocks or smaller.
You cannot access the data that is stored in the field over 60 TB. The external volumes of about 47 MB (96,000 Blocks) or smaller cannot be mapped.

(1) Register to an external volume group
When you map an external volume as an internal volume, you need to register the external volume to an external volume group.
The user can classify the external volumes, which is set by Universal Volume Manager, into the groups according to the use. The group is called external volume group (ExG). For instance, you can register multiple volumes in one external storage system to one external volume group. Or, even though the data you want to manage in a lump is stored in volumes in the different external storage systems, you can register the volumes in one external volume group and manage them in block.
You need to assign numbers from 1 to 16384 to external LU groups. A maximum of 14,080 external volume groups can be created. A maximum number of volumes, which can be registered in one external group, is 256.

Copyright © 2012, Hitachi, Ltd.

(2) Select the external volume attribute
When you map an external volume as an internal volume, you set the attributes of the external volume. The attributes of an external volume can be set using Add LU panel of Universal Volume Manager.

(a) I/O cache mode (Cache mode: Disable or Enable)
You can set if the operation uses the cache or not when the host I/O is demanded. If you select Enable, the host I/O once goes to the cache and then to the volume. If you select Disable, the host I/O comes directly to the volume not coming though the cache.

(3) Emulation type
You can set the emulation type of the mapped volume.

Copyright © 2012, Hitachi, Ltd.

3.15.2.6 Define LU paths

You can define LU paths of the external volumes.

3.15.2.7 Other settings

(1) Define alternate path to the external volume
When you map an external volume as an internal volume, the path(s) will be set from the internal volume to the external volume. When two paths, which can be set, from the two different clusters to the external volume exist, the two paths are set at the time of the mapping. When two settable paths do not exist, one path is set at the time of the mapping.
You can set up to eight paths to the external volume including the automatically set paths. Among the paths to the external volume, a path that is given the highest priority is called a primary path. Paths other than the primary path are also called alternative paths.
When the external volume is mapped as the internal volume using Universal Volume Manager, the host I/O operations to the external volume are enabled normally using the path set in the mapping operation. However, the path is automatically switched to the alternate path when the path set in mapping operation cannot be used due to, for instance, maintenance operation in the storage system, or a failure in the channel processor. Because the path is switched to the alternate path, you can continue performing the I/O operation to the external volume that is mapped by Universal Volume Manager as usual even though an error occurred in the original path.
If the alternate path is not set, host I/O, is aborted when a maintenance operation is performed for the storage system or a trouble such as a failure in the channel processor occurs.
It is recommended to set the alternate paths for safer operation.
To set the alternate path, use the path setting function of Universal Volume Manager.

 Copyright © 2012, Hitachi, Ltd.

Figure 3.15.2.7-1 illustrates an example of setting an alternate path. In Figure 3.15.2.7-1, external storage system ports, “WWN A” and “WWN B”, are connected to “CL1-A” and “CL2-A” respectively which are set to the external ports in the DW700 storage system. You need to specify the port of a different cluster in the DW700 storage system for the alternate path as “CL1” port and “CL2” port are specified in Figure 3.15.2.7-1.

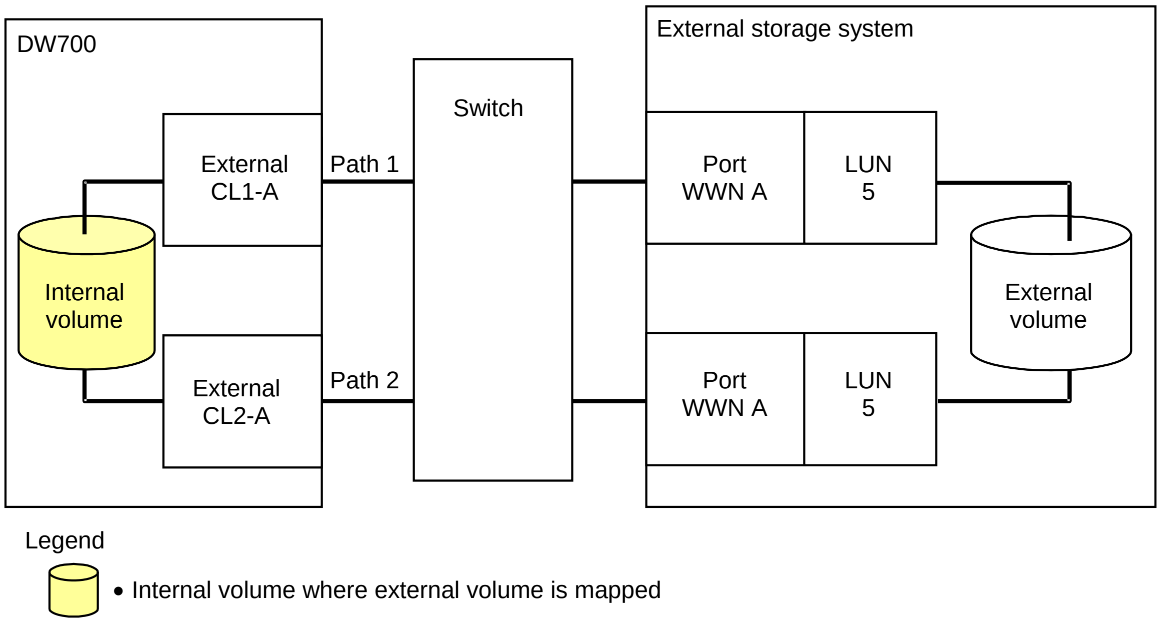


Figure 3.15.2.7-1

Figure 3.15.2.7-2 also illustrates an example of setting an alternate path. In Figure 3.15.2.7-2, two ports are specified in the DW700 storage system, and connected to the ports in the external storage system via the switch. In this case, two ports of different clusters are specified in the DW700 storage system. Therefore, the setting of the alternate path is enabled.
In Figure 3.15.2.7-3, two paths are also set between the internal volume and the external volume. However, one port is specified in the DW700 storage system, and two ports are specified in the external storage systems via the switch. Since two ports of different clusters need to be set in the DW700 storage system for alternate path settings in Universal Volume Manager, we do not recommend the setting shown in Figure 3.15.2.7-3.

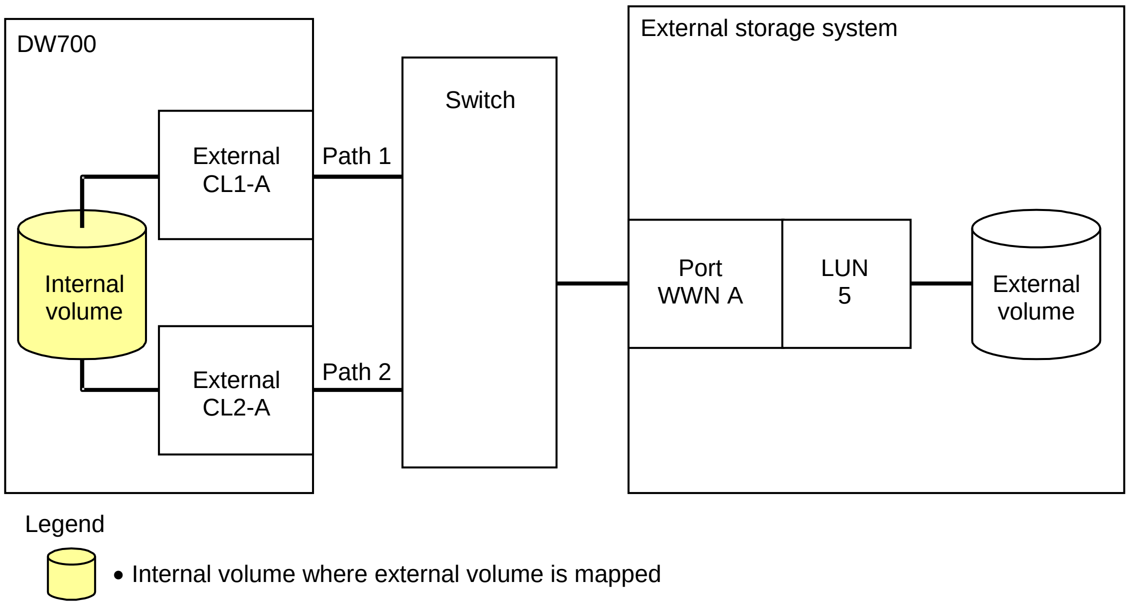


Figure 3.15.2.7-2

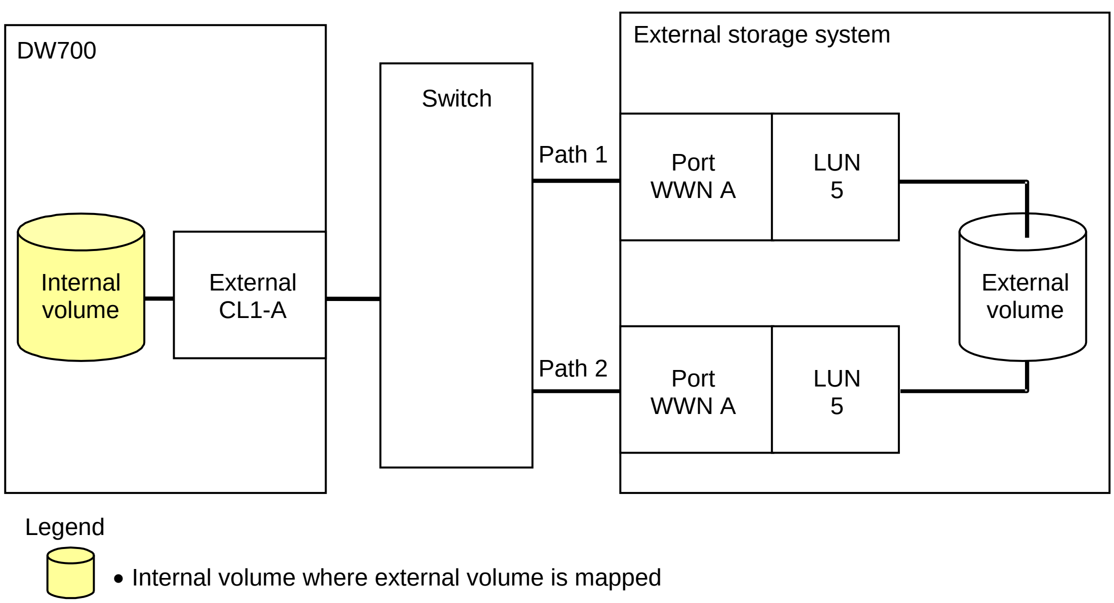


Figure 3.15.2.7-3

Copyright © 2012, Hitachi, Ltd.

3.16 (Blank)

Copyright © 2012, 2013, Hitachi, Ltd.

3.17 Cautions when Stopping the storage system

Pay attention to the following instructions concerning the operation when the storage system stops immediately after turning on the PDU breaker and completing the power off of the storage system.

3.17.1 Precautions in a Power Off Mode

Even if the PDU breaker is immediately after turning on, or the equipment is in the state of the power off after the process of the power off is completed, the equipment is in a standby mode. In this standby mode, the AC input is supplied to the power supply in the storage system, and this supplies the power supply to the FANs, the cache memory modules and some blades (MPBs, MAIN Blades, ENCs, or the like).
Please execute the following process when the standby electricity must be controlled, because the standby electricity (*1) exists in the equipment under this condition.

(1) When the equipment is powered on
Turn on the breaker of each PDU just before the power on processing.

(2) When the equipment is powered off
After the power off processing is completed, turn each PDU breaker off.
When turning each PDU breaker off, make sure the power off processing is completed in advance. When the breaker is turned off during the power off processing, the battery power is burned and the power on time may take little longer at the next power on processing depending on to the battery charge, because the equipment shifts to the emergency SSD transfer processing of the data with the battery. Moreover, in advance of the turning each PDU breaker off, make sure the AC cables of the other equipment are not connected to the PDU that is turned off.
The management information stored in the memory is transferred to the nonvolatile memory (SSD) in the MAIN Blade during the power off process, therefore, leaving the breaker of the PDUs on is needless for the data retention.

*1: Refer to Table 3.17.1-1 for standby electricity.

Table 3.17.1-1 Maximum Standby Electricity per Controller Chassis and Drive Box

Chassis/Box	Maximum Standby Electricity [VA]
CBX (Controller Chassis)	200
DBS (SFF Drive Box)	140
DBL (LFF Drive Box)	90
DBX (Drive Box (Dense))	420
DBF (Flash Module Drive Box)	320

Copyright © 2012, Hitachi, Ltd.

3.17.2 Operations when a distribution panel breaker is turned off

When the distribution panel breaker or the PDU breaker is turned off, execute the operation after making sure that the power supply of the storage system was normally turned off, by confirming the PS was normally turned off according to the Power OFF Event Log (Refer to 3.9.2 Power OFF Procedure in INSTALLATION SECTION.) of SVP.

When the Power OFF Event Log confirmation is impossible, suspend the operation and request customers to restart the storage system to confirm that the PS is normally turned off.

NOTICE: Request the thoroughness in the following operations to the customer if the distribution panel breaker or the PDU is impossible to keep the on-state after the power of the storage system is turned off.

1. Point to be checked before turning the breaker off
   Request the customer to make sure that the power off processing of the storage system is completed (READY lamp, MESSAGE lamp and ALARM lamp are turned off) before turning off the breaker.

2. Operation when breaker is turned off for two weeks or more
   A built-in battery spontaneously becomes a state of discharge when the distribution panel breaker or PDU breaker is turned off after the storage system is powered off. Therefore, when the breaker is turned off for two weeks or more, charging the built-in battery to full will take a maximum of 4.5 hours. Accordingly, request the customer to charge the battery prior to the restarting of the storage system.

Copyright © 2012, Hitachi, Ltd.

3.18 System Disk

The system disk is a volume for storing information used by the storage system. The overview of the system disk is shown below.

3.18.1 Setting the System Disk

The Virtual LVI/LUN software is required for setting the system disk. Set the system disk using the Install CV operation or Make Volume operation of this software.

The use and required capacity of the system disk are shown in the table below.

Use	In case of OPEN-V
Buffer area of Audit log	130 Mbytes

When using the system disk for the above use, prepare the system disk with the above-mentioned required capacity remained in one volume. For example, if you set one volume of the system disk of 200 Mbytes, it is usable for storing for buffer area of the Audit log. However, even if you set two volumes of 100 Mbytes, you cannot use them for storing buffer area of the Audit log.

Copyright © 2012, Hitachi, Ltd.

The cautions about setting the system disk are as shown below.

- Do not store the usual data in the system disk.
- In case of the mixed configuration of the internal and external volumes, define the system disk as the internal volume.
- Do not perform the BIND setting of DCR to the system disk.
- When two or more system disks are defined, it is used for order with a young LDEC:CU:LDEV number.
- When two or more system disks are defined, other system disks are not used by the automatic operation at the failure of the system disk.
- When Mode676 is turned on for the Buffer area of Audit log, the use of the system disk becomes effective.

 Copyright © 2012, Hitachi, Ltd.

3.18.2 Protecting the System Disk

The system disk is protected from the de-installation/blockade operation, host I/O and maintenance operation as shown below.

The de-installation of the system disk is performed using the Volume To Space operation. However, the de-installation/blockade operation may be denied as shown in the table below depending on the use status of the system disk.

Case to be denied	Actions to be taken
System disk is being used for Audit log.	Execute it again after release the MODE676.
Blocked system disk	Execute it again after changing to the normal status. (If you cannot change it, you can execute it by entering the password.)

The system disk is protected from open I/O by denying the SCSI path definition for the system disk.

The following maintenance operation for the system disk cannot be performed.

- Creating pair of TrueCopy/UR/ShadowImage/Thin Image
- Instructing VOL migration of Volume Migration and setting reserve VOL
- Setting access right of Volume Security
- Setting the attribution other than Read/Write possible of Data Retention Utility and Volume Retention Manager
- Setting the journal volume of UR
- Instructing LDEV connection by LUSE
- Setting PAV
- Volume Initialize of CVS
- Setting the SCSI path definition
- Setting the command device/remote command device
- Setting Write Through suppression setting VOL/pool-VOL

Copyright © 2012, Hitachi, Ltd.

3.18.3 Failure recovery of the System Disk

The failure recovery procedure of the system disk is shown in the table below.

Use	Failure recovery procedure
Buffer area of Audit log	① Release the MODE676. ② De-install the system disk using the Volume To Space operation. ③ When the hardware has a problem, perform the maintenance action based on the action code instruction. ④ Reset the system disk, which is once de-installed on the process ②, with the Install CV or the Make Volume operation. ⑤ Set the MODE676.

Copyright © 2012, Hitachi, Ltd.

3.19 CVS and DCR Option function

3.19.1 Customized Volume Size (CVS) Option

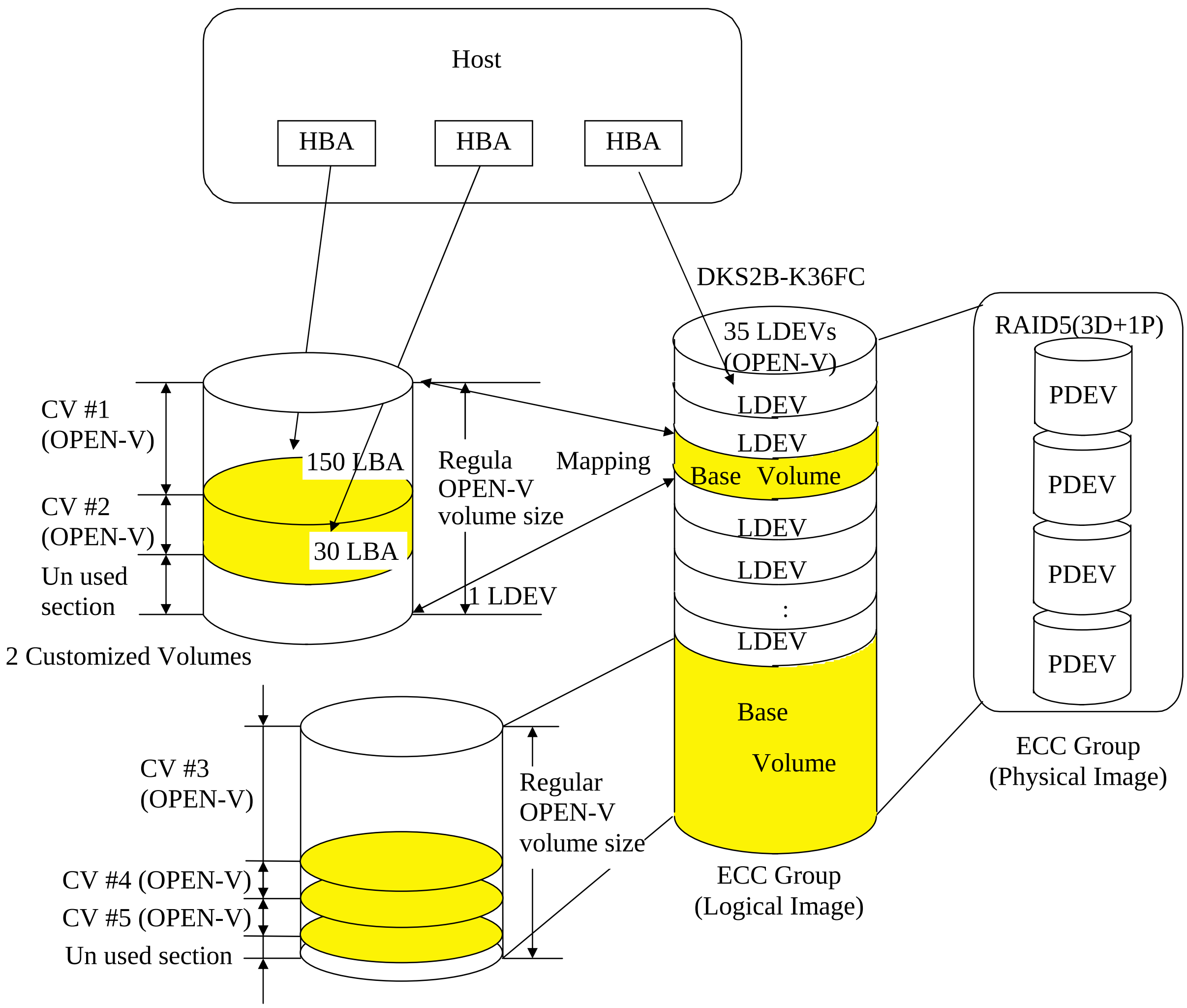
3.19.1.1 Outline

When two or more files to which I/Os are applied frequently exist in the same volume, a contention for the logical volume occurs. If this occurs, the files mentioned above are stored separately in different logical volumes and an action is taken to avoid contention for access. (Or means to prevent I/Os from generation is required.)
However, the work for adjusting the file arrangement giving consideration to the accessing characteristic of the file will be a burden on users of the DKC and it is not welcomed by them.
To solve this problem, the Customized Volume Size (CVS) option is provided. (Hereinafter, it is abbreviated to CVS.)
The CVS provides a function for freely defining the logical volume size. By doing this, even in a storage system with the same capacity, the number of volumes can be increased easily. As a result, a file with a high I/O frequency can be easily allocated to an independent volume. That is to say, the trouble to consider a combination of stored files in a volume can be saved.

 Copyright © 2012, Hitachi, Ltd.

3.19.1.2 Features

- The capacity of the ECC group can be fully used.
- By combining with the Dynamic Cache Residence (DCR) option, high performance equivalent to that of a semiconductor storage device can be realized.



 Copyright © 2012, 2013, Hitachi, Ltd.

3.19.1.3 Specifications

The CVS option consists of a function to provide variable capacity volumes.

(a) Function to provide variable capacity volumes
This function can create the capacity volume as required by the users.
You can set the data by Mbytes or Logical Blocks.

Table 3.19.1.3-1 CVS Specifications

Parameter	Content
Track format	OPEN-V
Emulation type	OPEN-V
Maximum number of volumes (normal and CVS) per VDEV	2,048 for RAID5 (7D+1P), RAID6 (6D+2P) or RAID6 (14D+2P) 1,024 for other RAID levels
Maximum number of volumes (normal and CVS) per storage system	16,384
Minimum size for one CVS Volume	48 MB
Maximum size for one CVS Volume	Internal volume: 3,221,159,680 KB (2.99 TB) External volume: 4,294,967,296 KB (4 TB)
Size increment	1 MB
Disk location for CVS Volume	Anywhere

Copyright © 2012, Hitachi, Ltd.

3.19.1.4 Maintenance functions

Features of the maintenance functions of the CVS option is that they allow execution of not only the conventional maintenance operations instructed by the SVP but also the maintenance operations instructed from the Storage Navigator. (See Item No. 2 to 5 in Table 3.19.1.4-1.)
Unlike the conventional LDEV addition or reduction, the operation for the ECC group is made unnecessary, so that the volumes can be operated from the Storage Navigator.

Table 3.19.1.4-1 Maintenance Function List

Item No.	Maintenance function	CE	User	Remarks
1	Concurrent addition or deletion of CVs at the time of addition or deinstallation of ECC group	✓	—	Same as the conventional addition or deinstallation of LDEVs
2	Addition of CVs only	✓	✓	Addition of CVs in the free area
3	Conversion of normal volumes to CV	✓	✓	
4	Conversion of CV to normal volumes	✓	✓	
5	Deletion of CVs only	✓	✓	Only the optional CVs are deleted and incorporated into the free area. No deinstallation of ECC group is involved.

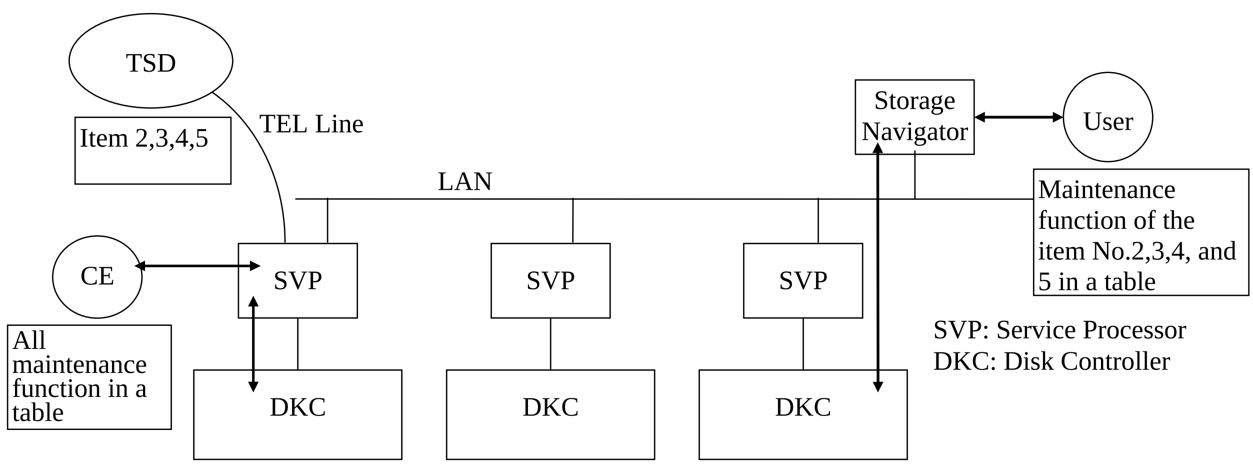


Fig. 3.19.1.4-1 Maintenance Execution Route when CVS Is Used

Copyright © 2012, Hitachi, Ltd.

3.19.2 Dynamic Cache Residence (DCR) Option

3.19.2.1 Outline

Because cache capacity is usually smaller than the total HDD capacity in RAID storage system, cache cannot be keep all data at the same time which is stored whole HDDs in the storage system. Cache management control cache extent so as to allocate more capacity to more frequently accessed data to embed this gap by LRU algorithm.
By this control, data with a low access frequency is hard to remain in the cache, de-staged into HDDs, therefore the more frequent accesses to physical HDDs occur, decrease the access performance and unpredictable response time appear.
The DCR option provides a function for making data being resident on the cache and realize high access performance.
DCR is a function to make the specified data resident in the specific Cache area in DKC. This will enable the host to always execute Cache hit of the specified data and access it at high speed.

Copyright © 2012, Hitachi, Ltd.

3.19.2.2 Features

- Feature of the DCR consist of two modes, one is called “PRIOrity mode(hereinafter,it is abbreviated to PRIO)” and the other “BIND mode(hereinafter,it is abbreviate to BIND)”.
  PRIO is a basic mode(100% Read Hit) of this feature, which fits typical user needs and BIND is supplementary for special customer(100% Read/Write Hit. Replace SSD) needs.

- To use the DCR, addition of the cache memory is required for the service as a “DCR cache”.

- DCR supports PreStaging function. The PreStaging is a function which read data on Logical Volume onto cache by receiving Storage Navigator or SVP instructions.

3.19.2.2.1 PRIO

[Processing]

- When a read command to the data which is assigned as DCR extent is issued and first meet "miss" in cache, the data is staged into cache by usual staging mechanism.
  The data remains in the DCR cache permanently for future access to guarantee read-hit performance even after data is transferred to the host, regardless usually cache LRU management.
  If a write command is issued to the data remaining in the cache in the way above, the data is updated with an out-of-DCR cache segment provided for write data duplication, and the data is de-staged into the HDD. Then the cache segment for write data duplication is returned to the out-of-DCR cache segment group.
  In this case, new data is left in the cache extent together with old data.

[Performance impact]

- Theoretically, because above de-staging into HDD is processed by usual asynchronous de-staging mechanism, succeeding host access has a possibility to meet the same cache slot collision by locking both the host access and asynchronous de-staging process.
  However in that case, by minimizing the collision time implemented in micro-code, performance impact will be negligible for usual customer jobs.
- In the case the storage system cache becomes overloaded, a performance degradation occurs because the non-DCR cache segment must be used for the data assurance during a period between write data reception and de-staging operation completion.

NOTICE:	When the operation of deleting the resident cache-data is performed during host I/O execution, the host I/O execution conflict with the procedure in which the data is transferred to disk drives (de-staging) may happen. It may cause the response performance degradation.

To avoid the response performance degradation, please limit the total capacity of data released by one operation.

- If the Host timeout period is more than 11 seconds, the amount of acceptable releasing cache is limited to 3Gbyte or less.
- If the Host timeout period is less than 10 seconds, it is limited to 1Gbyte or less.

NOTICE:	If the setting or the release of the DCR extent is performed to a lot of LDEVs when there is I/O from Host, the response performance degradation of HOST I/O may occur. To avoid the response performance degradation, limit the number of LDEVs to be set or released by one operation to 1.

Copyright © 2012, Hitachi, Ltd.

[Maximum DCR capacity]

- Addition of the “DCR cache” is required for the DCR, and a number of disk tracks equivalent to the capacity of the added “DCR cache” is the maximum number definable for the DCR.
  Besides, STR recommends to keep the standard cache capacity for the non-DCR portion as a minimum out-of-DCR cache, to avoid considerable performance degradation for original data because of the newly installed DCR (Standard Cache capacity is decided by the storage system capacity).
  To keep this rule, when the customer want to install DCR feature for the storage system, he needs to install additional cache capacity as a DCR area is taken out of pre-defined standard cache capacity.
  Table 3.19.2.2.1-1 indicates the additional cache capacity as a DCR area out of pre-defined standard cache capacity.
  The additional cache capacity requires out-of-DCR cache capacity.
  Table 3.19.2.2.1-1 is the relationship between number of using extents and additional cache capacity.

Copyright © 2012, Hitachi, Ltd.

Table 3.19.2.2.1-1 Necessary addition of cache memory

	Number of Priority Mode extents			
	1 ~ 4096	4097 ~ 8192	8193 ~ 12288	12289 ~ 16384
Additional Standard cache memory capacity	16GB	16GB	32GB	32GB

*1: 1GB=1,024^3 Byte

Caution

A required cache capacity in PRIO mode:
standard cache capacity + DCR cache capacity + the above cache capacity

A required cache capacity in BIND mode:
standard cache capacity + DCR cache capacity

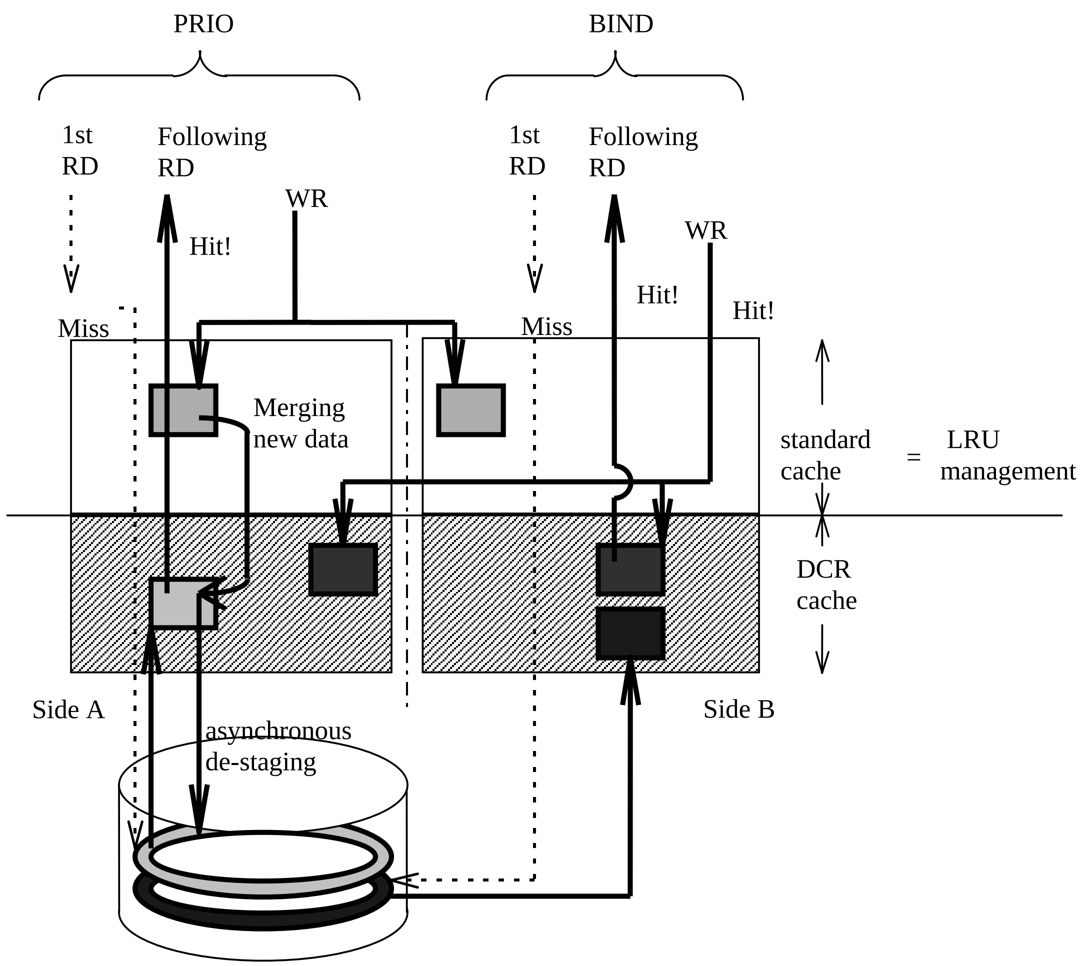


Fig. 3.19.2.2.1-1 Processing for DCR Extent

3.19.2.2.2 BIND

[Processing]

- As described above, there is a possibility that a responding performance degradation occurs in the PRIO mode because of a collide slot lock between the host access and asynchronous de-staging process or a waiting for the cache segment to be empty caused by an overload of the storage system cache.
  This is a negligibly small impact on the performance in the typical user environment, however, in some environments, if any, in which the performance is very critical and the maximum response time must be assured, the above factors to degrade the response may be the issues.

[Performance impact]

- In BIND, the difference for PRIO, not only read data but also being all write data for the assigned DCR extent are resident in cache, no de-stage process occurred by any write command. Thus, by protecting any asynchronous de-stage process for the DCR data, read operation become a perfectly hit process.
- However, as a compensation for the perfect hit performance, the cache, which is three times larger than that in the PRIO mode, is used.

NOTICE:	When the operation of deleting the resident cache-data is performed during host I/O execution, the host I/O execution conflict with the procedure in which the data is transferred to disk drives (de-staging) may happen. It may cause the response performance degradation.

To avoid the response performance degradation, please limit the total capacity of data released by one operation.

- If the Host timeout period is more than 11 seconds, the amount of acceptable releasing cache is limited to 3Gbyte or less.
- If the Host timeout period is less than 10 seconds, it is limited to 1Gbyte or less.

NOTICE:	If the setting or the release of the DCR extent is performed to a lot of LDEVs when there is I/O from Host, the response performance degradation of HOST I/O may occur. To avoid the response performance degradation, limit the number of LDEVs to be set or released by one operation to 1.

[Maximum DCR capacity]

- User data can be set 1/3 of cache capacity for BIND.

Copyright © 2012, 2013, Hitachi, Ltd.

3.19.2.2.3 Assignment of DCR extent and guard logic

- BIND/PRIO modes can be assigned for each DCR extent individually.
  For example, the user want to assign 1GB cache for PRIO and 256MB cache for BIND, DKC allocate the extents in additional cache area by 1GB for PRIO and 256MB times 3 equals 0.75GB for BIND. Total 1.75GB cache should be additionally installed. The user can assign repeatedly many DCR extents with choosing each mode for each extent. Say, 1GB for PRIO + 512MB for PRIO + 256MB for BIND + 128MB for BIND comes to 2.6GB total DCR area necessary. The real capacity of cache needs adjusting to the cache unit to be added.
- SVP micro-code accepts many extents allocated for DCR. A minimum of 4GB must remain unallocated to DCR.
  In other words, DKC checks that the remaining cache capacity after DCR allocation is over 4GB, if the addition of DCR breaks the remaining 4GB boundary, the SVP rejects the allocation of DCR extents with an error message.
- From that point, the user theoretically can assign all cache capacity except 4GB for DCR regardless of their configurations, STR strongly recommend the user should keep standard cache capacity out of DCR dependant on the configuration according to the manuals. Guard boundary by SVP is only 4GB for all configurations.
- For the outline and the setting operation procedure of the DCR cache, see page THEORY03-19-190.

Copyright © 2012, Hitachi, Ltd.

3.19.2.2.4 DCR PreStaging

- The processor reports SIMs when the PreStaging abnormal end.

(following Table)

Error		REF CODE			SIM	Level	Host report	Remarks
		22	23	13	28			
DCR status	PreStaging abnormal end	48	21	xx	FE	Service	No	xx: abnormal end reason code

x‘E1’, x‘10’ : No DCR PP
x‘E2’, x‘20’ : Storage system busy
x‘E4’, x‘40’ : Staging time over
x‘E5’, x‘50’ : Cache or SM blockade
x‘E6’, x‘60’ : LDEV warning
x‘E7’, x‘70’ : Staging failure
x‘E8’, x‘80’ : P/S OFF
x‘E9’ x‘90’ : PreStaging canceled
x‘EA’, x‘a0’ : Cache over loaded
x‘EB’, x‘b0’ : Some MP’s blockade

- In the case the storage system cache becomes overloaded, a resulting performance degradation occurs during a PreStaging execution. We strongly recommend issuing a PreStaging request to stage data onto cache at the timing of normal load, or SIM REF CODE = 4821a0 may be reported resulting in failure.
- DKC rejects PreStaging requests during PreStaging execution. Please retry PreStaging requests after PreStaging termination.

If you specify the DCR setting on the volume during the quick formatting, do not use the prestaging function. If you want to use the prestaging function after the quick formatting processing completes, first you need to release the setting and then specify the DCR setting again, with the prestaging setting enabled this time. For information about the quick formatting, see the “Provisioning Guide”.

Copyright © 2012, Hitachi, Ltd.

3.19.2.3 Specifications

Table 3.19.2.3-1 Specifications of the Function

Item No.	Item	Description
1	Maximum number of areas to be made resident	For the PRIO and BIND modes together: 4096 areas/logical volume 16384 areas/storage system
2	Unit of area specified to be resident	512 logical blocks
3	Minimum/Maximum size of extent	512 logical blocks/logical volume size
4	Online change of resident area	Allowable (from the SVP and Web Console)
5	Addition of cache capacity	Mandatory(Program Product: Charged with cache)
6	Maximum usable cache capacity as DCR	Capacity of the cache memory added as the DCR cache. The “standard cache capacity” must be ensured by the rule.

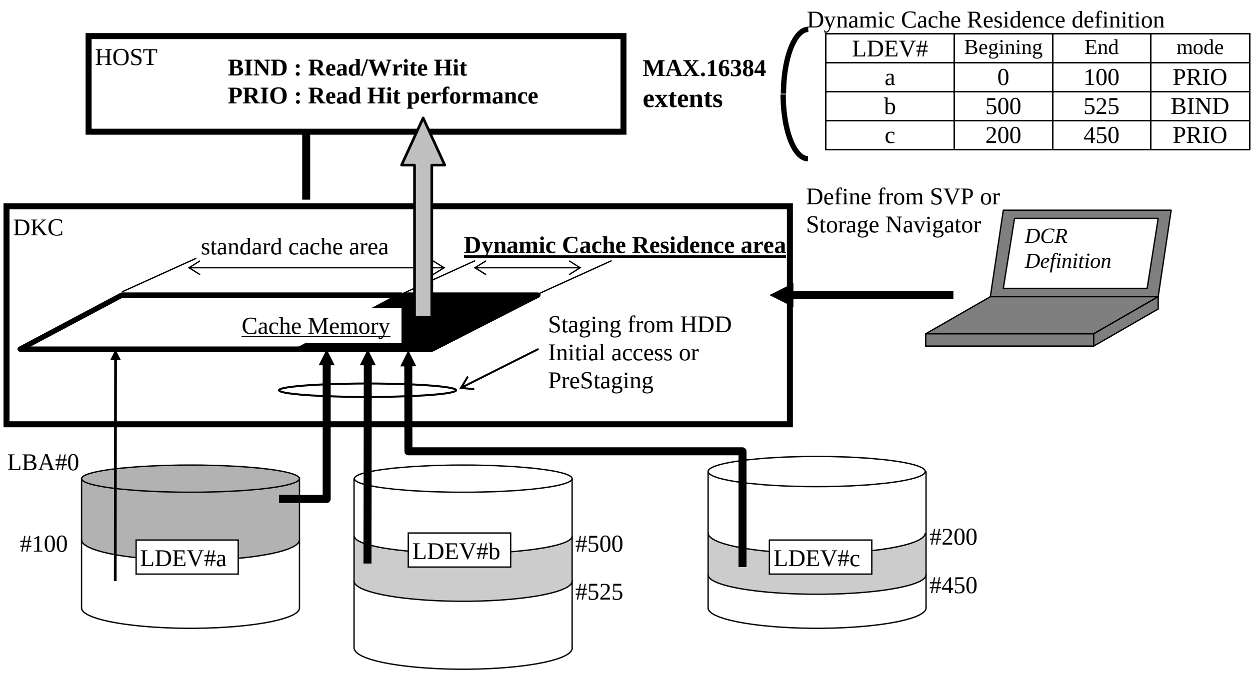


Fig. 3.19.2.3-1 Conceptual Diagram of DCR

 Copyright © 2012, Hitachi, Ltd.

3.19.2.4 Maintenance functions

The characteristic of the maintenance function of the DCR option is that the maintenance operation is possible from not only the SVP instruction but also Storage Navigator. (Refer to Table 3.19.2.4-1.)

Table 3.19.2.4-1 Maintenance Function List

Item No.	Maintenance operation	TSD	CE	User	Description
1	Addition of the DCR area	—	✓	✓	• Adds the DCR area. • Unit of the area to be specified by the SVP is a track.
2	Deletion of the DCR area	—	✓	✓	• Deletes the continuous DCR area.
3	Change of the DCR area	—	✓	✓	• Changes the DCR area size.
4	Status display of the DCR area	—	✓	✓	• Displays the specifications of the DCR area.
5	Addition and de-installation of the DCR cache	—	✓	—	• Because insertion and pulling off of the cache module into/from the DKC is required
6	Indication of DCR PreStaging	—	✓	✓	• Indicates the DCR PreStaging.

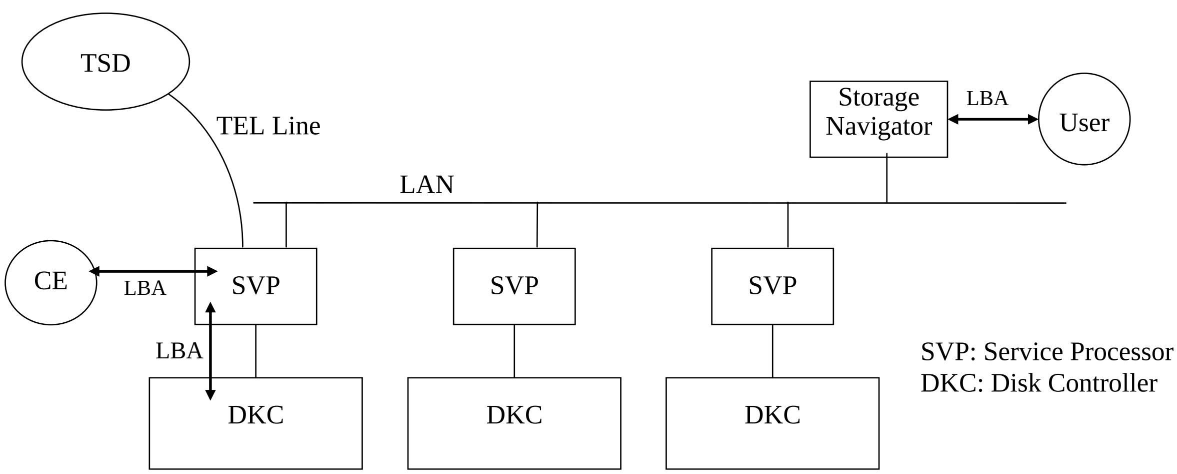


Fig. 3.19.2.4-1 Maintenance Execution Route when DCR Is Used

Copyright © 2012, Hitachi, Ltd.

3.19.2.5 Notes on maintenance when DCR is used

When performing the following maintenance, it is attended with a temporary regression of the cache memory or shared memory. Since this regression disables retention of the cache capacity required for the DCR, the DCR function is automatically suppressed for a while until the maintenance is completed.
A service person is required to obtain an approval of a user before starting the maintenance because there is a high possibility that the DCR function suppression may result in a degradation of responding performance.

(1) Cache replacement
(2) Addition and de-installation of the cache (including addition and de-installation of the DCR cache)
(3) Addition of the ECC Gr. and LDEV requiring addition of the SM because of the addition of the CU
(4) Addition and de-installation of the SM
(5) Cluster maintenance

We recommend that CE or Customer should execute the following action after DKC power supply restoration or equipment restoration, when the DKC power supply is down by power failure or mistake during “DCR area” release.
(Because it is a high possibility that the action may result in a degradation of responding performance, CE should execute the following action on a customer’s authority.)

Action : (1) CE or Customer should release all DCR areas in a DCR area released Volume.
(2) CE or Customer should set up again all DCR areas with the exception of the released DCR area in the DCR area released Volume.

Reason : When DKC power is off during “DCR area” release, it is possible that DKC left a release DCR data on Cache.
DKC does not faulty operation by leaving “released DCR data” use a excessive cache memory.
We recommend that CE or Customer should execute the above-mentioned action after DKC restoration, because DKC perfectly execute the “DCR area” release process.

Copyright © 2012, Hitachi, Ltd.

3.19.2.6 Effects of DKC failures on DCR

The DCR function is automatically suppressed when any of the following failures occurs. The suppression continues until the regressed operation owning to the maintenance is canceled in the cases of Items (1) to (3) or continues until an automatic recovery of the shared memory by the microprogram is terminated normally in the case of Item (4).

(1) Cache failure

(2) Shared memory failure

(3) One-side cluster down

(4) One-side shared memory blockade (SIMRC = FFEE)

Copyright © 2012, Hitachi, Ltd.

3.19.2.7 Automatic cancellation of DCR

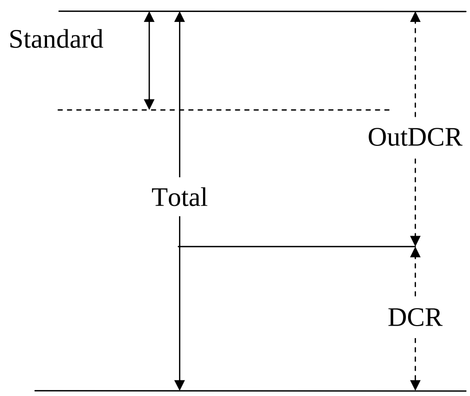
The DCR setting of a volume to be de-installed by the functions of Deletion of CVs (LDEV), Conversion of CV to normal volume, and Conversion of normal volume to CV is automatically canceled as a part of the de-installation processing by the SVP microprogram.

Copyright © 2012, Hitachi, Ltd.

3.19.2.8 Explanation of DCR cache and procedure for setting operation

3.19.2.8.1 Explanation

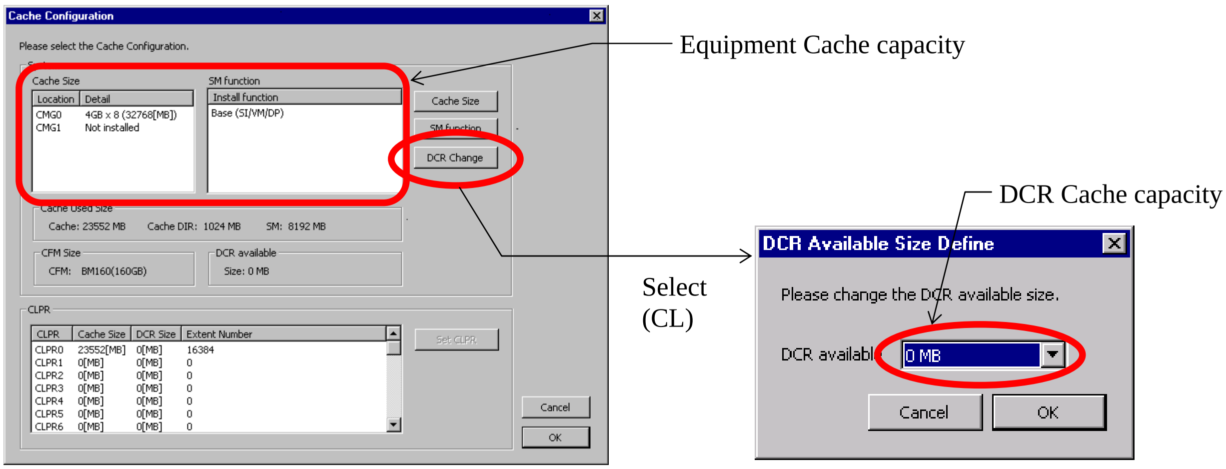
- A cache module must be defined and installed for the DCR before using the DCR.
- The DCR extent can be set only for the defined "DCR cache capacity."
- The "out-of-DCR cache capacity" must be retained more than "standard cache capacity" which is defined in accordance with the disk capacity in order to assure the performance in the non-DCR area.
- Therefore, DCR extent definition more than "DCR cache capacity" is rejected according to the SVP guarding logic. Also, the "DCR cache capacity" definition lower than the minimum "standard cache capacity" (2GB × 2) is also rejected.



Total : Equipment Cache capacity
OutDCR : Out-Of-DCR Cache capacity
DCR : DCR cache capacity
Standard : Standard Cache capacity

3.19.2.8.2 Setting operation procedure

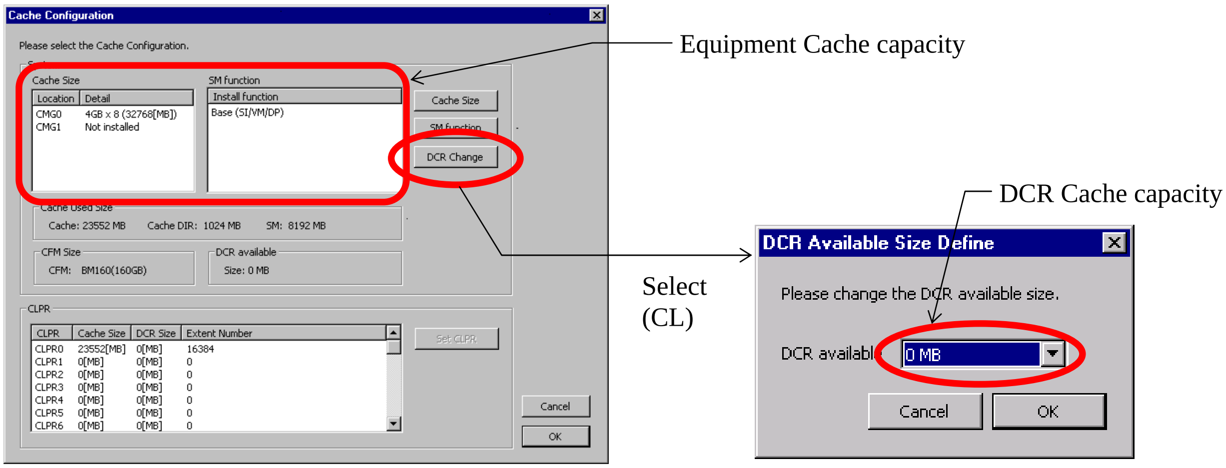
(1) Setting DCR cache capacity in Define Config & Install sequence
Set the DCR cache capacity in the equipment cache capacity.



NOTE: Set the DCR cache capacity so that it is less than the “equipment cache capacity minus standard cache capacity.”

Copyright © 2012, Hitachi, Ltd.

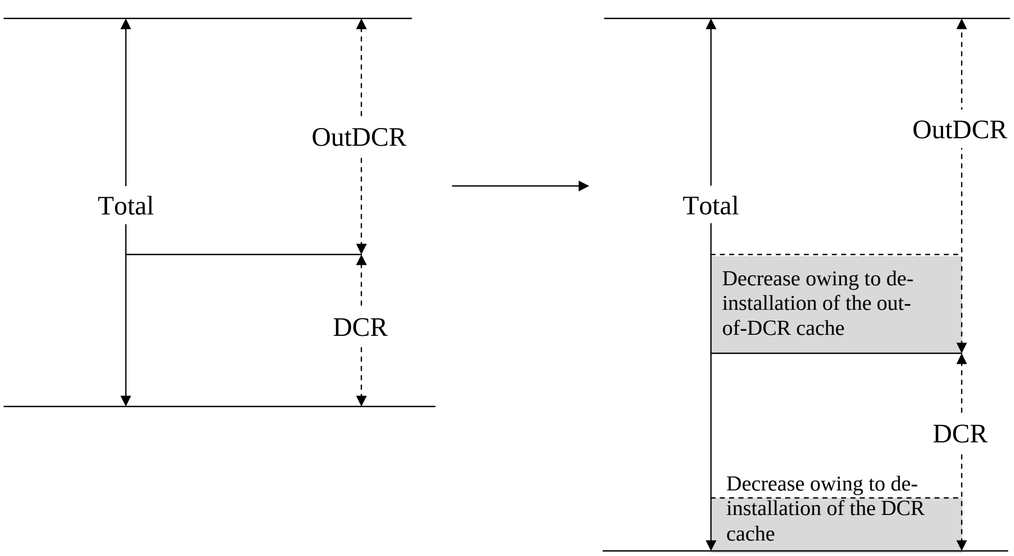
(2) Adding operation of DCR cache capacity in cache addition sequence
When adding the cache, set the DCR cache capacity in the equipment cache capacity after the addition.



For example, to change a status with the cache of 2.0 GB × 2 installed and 256 MB × 2 of it set to the DCR cache to a status with the cache of 4.0 GB × 2 installed and 512 MB × 2 of it set to the DCR cache by adding the cache of 2.0 GB × 2,

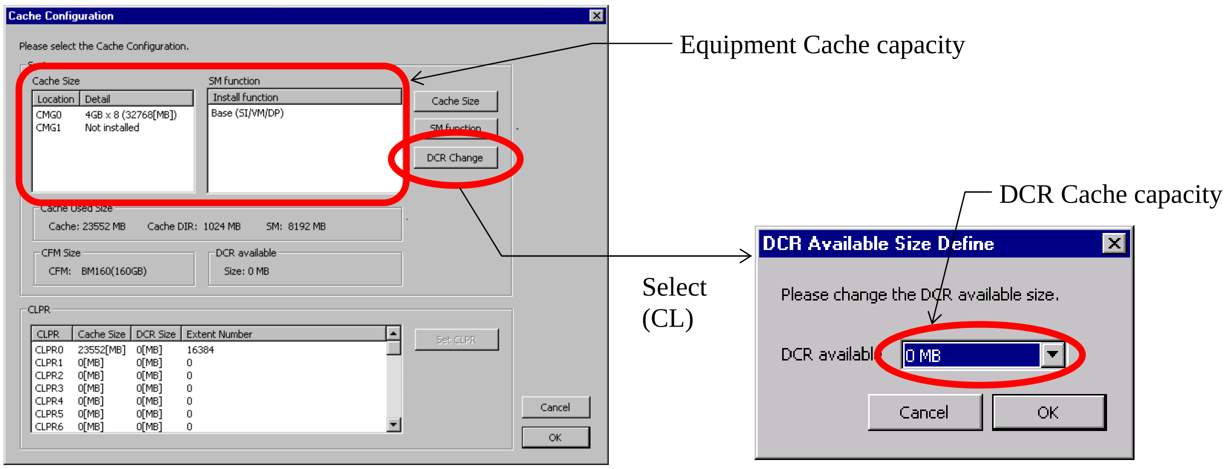
① set the equipment cache capacity to 4.0 GB × 2 in the "Cache Configuration" dialog box, and
② press the "Change..." button to open the "DCR Available Size" dialog box and set the DCR cache capacity to 512 MB × 2.

The DCR cache capacity can be set up to the cache capacity to be added. In the above example, the DCR cache capacity can be set up to 768 MB × 2 by adding 512 MB × 2.



Copyright © 2012, Hitachi, Ltd.

When de-installing the cache, set a capacity to be left as the DCR cache in the equipment cache capacity.



For example, to change a status with the cache of 4.0 GB × 2 installed and 512 MB × 2 of it set to the DCR cache to a status with the cache of 2.0 GB × 2 installed and 256 MB × 2 of it set to the DCR cache by de-installing the cache of 2.0 GB × 2,

① set the equipment cache capacity to 2.0 GB × 2 in the “Cache Configuration” dialog box, and

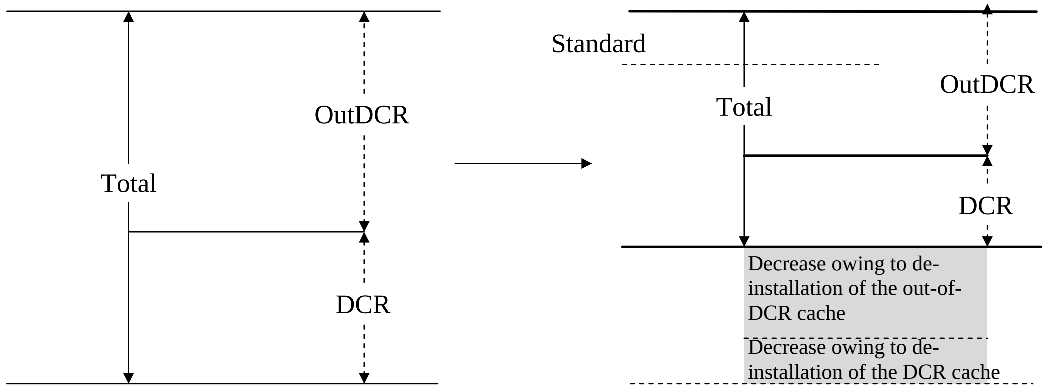
② Press the “Change...” button to open the “DCR Available Size” dialog box and set the DCR cache capacity to 256 MB × 2.

The maximum decreasable capacity of the DCR cache is equal to the capacity of the installed cache to be de-installed. The maximum decreasable capacity of the DCR cache capacity in the above example is 2.0 GB × 2. As a result, the DCR cache capacity after the de-installation becomes 0 MB × 2.

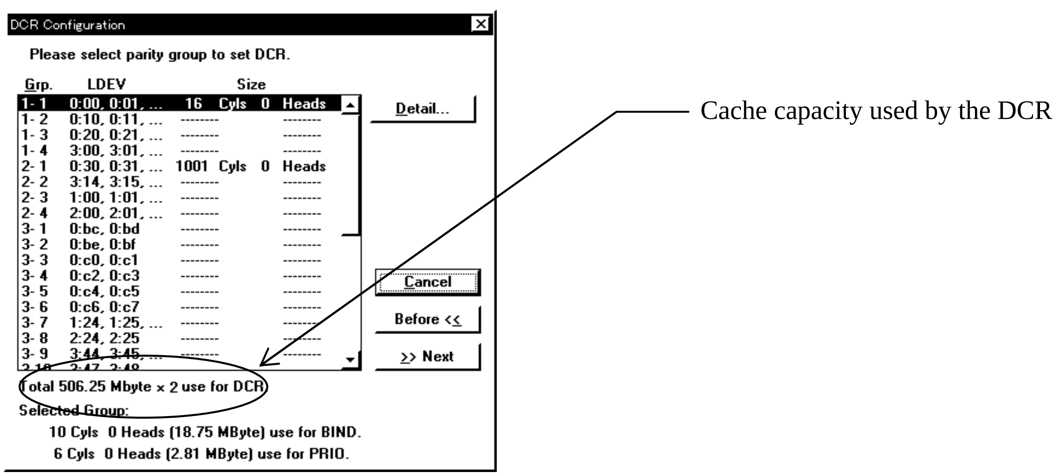
NOTE:

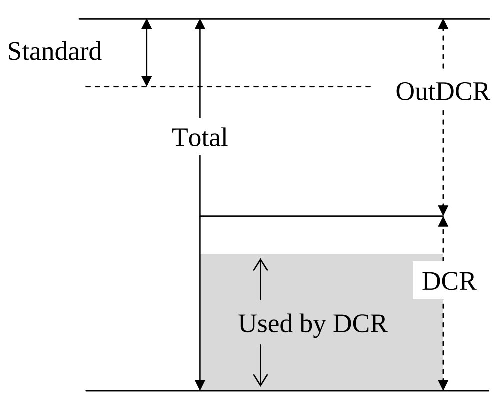
- In the case in which the de-installation of the DCR cache causes the capacity used by the DCR actually defined as the DCR extent to be above the DCR cache capacity, the cache de-installing process is suspended by the SVP guarding logic. Before executing the DCR cache de-installation, cancel the DCR setting to decrease the actual capacity used by the DCR.
- It is required to avoid de-installation of the out-of-DCR cache which causes its capacity to be below the standard cache capacity.

 Copyright © 2012, Hitachi, Ltd.



- The “cache capacity used by the DCR” actually used as the DCR extent is displayed on the DCR Configuration screen in [SVP] - [Install] - [Refer Configuration] for confirmation.





Copyright © 2012, Hitachi, Ltd.

3.19.2.8.3 Notes at the time of operation

From a cash manager utility, while carrying out DCR setup / release, please do not perform DCR operation (a display is included) of SVP/Web Console simultaneously.

When the Volume is in quick formatting, please do not set and release DCR from Cache Manager Utility until the operation is completed. For information about the quick formatting, see “Provisioning Guide”.

 Copyright © 2012, 2013, Hitachi, Ltd.

3.20 Caution of Flash Drive and Flash Module Drive installation

For caution of Flash Drive and Flash Module Drive installation, refer to INST01-130.

Copyright © 2012, Hitachi, Ltd.

3.21 Data guarantee

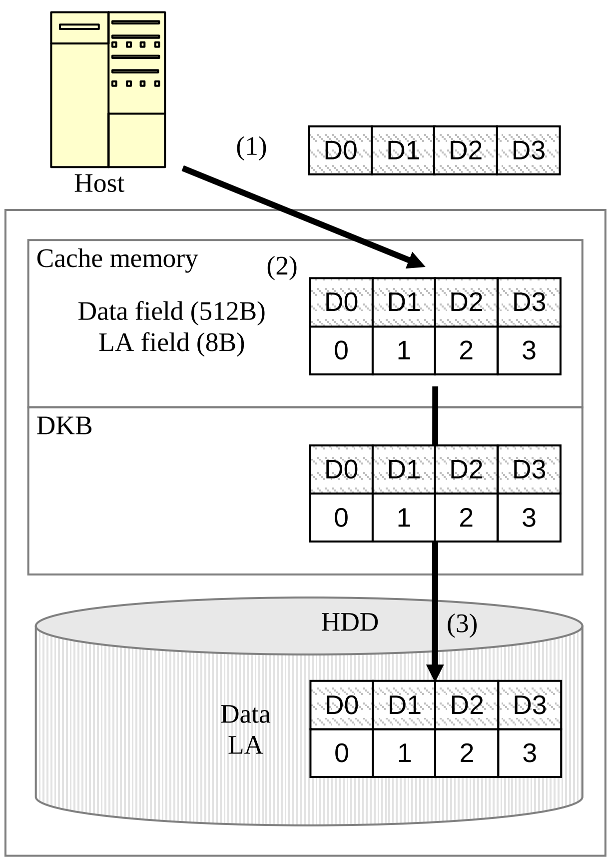
DW700 makes unique reliability improvements and performs unique preventive maintenance.

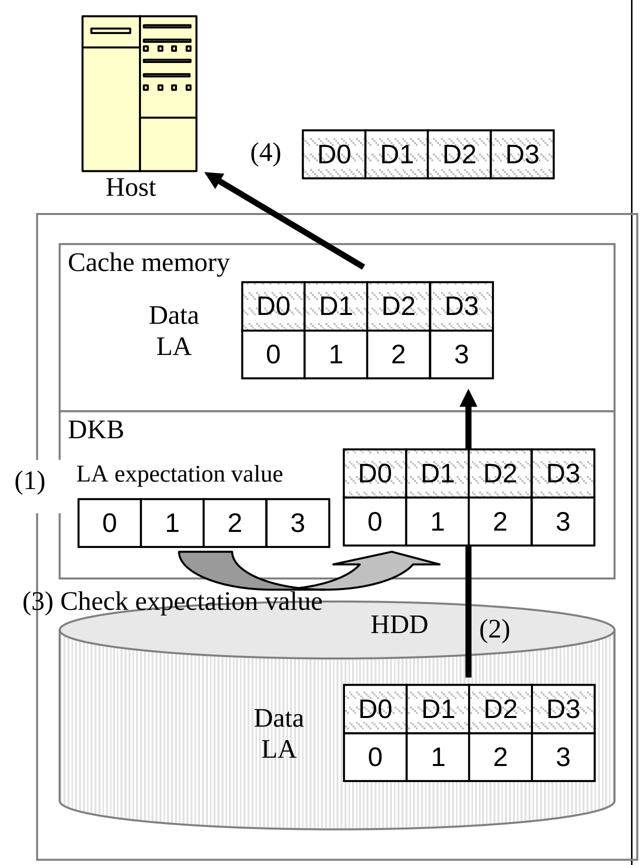
Copyright © 2012, Hitachi, Ltd.

3.21.1 Data check using LA (Logical Address) (LA check) (Common to SAS drives and SSD)

When data is transferred, the LA value of the target BLK (LA expectation value) and the LA value of the actual transferred data (Read LA value) are compared to guarantee data. This data guarantee is called “LA check”. With the LA check, it is possible to check whether data is read from the correct BLK location.

Table 3.21.1-1 LA check method

<table>
<tr><th>Write</th><th>Read</th></tr>
<tr><td>


</td><td>


</td></tr>
<tr><td>
(1) Receive Write request from Host.
(2) CHB stores data on cache and adds the LA value and, at the same time, adds an LA value, which is a check code, to each BLK. (LA value is calculated based on the logical address of each BLK)
(3) DKB stores data on HDD.
</td><td>
(1) DKB calculates the LA expectation value based on the logical address of the BLK to read.
(2) Perform read from HDD.
(3) Check whether the LA expectation value and the LA value of the read data are consistent. (When the LBA to read is wrong, the LA values would be inconsistent, and the error can be detected. In such a case, a correction read is performed to restore data.)
(4) CHB transfers data to Host by removing the LA field.
</td></tr>
</table>

Copyright © 2012, Hitachi, Ltd.

3.22 PDEV Erase

3.22.1 PDEV Erase

3.22.1.1 Overview

When the specified system option (*1) is set, the DKC deletes the data of PDEV automatically in the case according Table 3.22.1.1-2.

*1: Please contact to T.S.D.

Table 3.22.1.1-1 Overview

No.	Item	Content
1	SVP Operation	Select system option from “Install”.
2	Status	DKC only reports on SIM of starting the function. The progress status is not displayed.
3	Result	DKC reports on SIM of normality or abnormal end.
4	Recovery procedure at failure	Re-Erase of PDEV that terminates abnormally is impossible. Please exchange it for new service parts.
5	P/S off or B/K off	The Erase processing fails. It doesn’t restart after P/S on.
6	How to stop the “PDEV Erase”	Please execute Replace from the Maintenance screen of the SVP operation, and exchange PDEV that Erase wants to stop for new service parts.

Table 3.22.1.1-2 PDEV Erase execution case

No.	Execution case
1	PDEV is blocked according to Drive Copy completion.

 Copyright © 2012, 2013, Hitachi, Ltd.

3.22.1.2 Rough estimate of Erase time

The Erase time is decided by capacity and the rotational speed of PDEV.
Time is indicated as follows. (Time is a standard and it might take the TOV)

Table 3.22.1.2-1 PDEV Erase completion expectation time (1/2)

No.	Kind of PDEV	200GB	300GB	400GB	600GB	800GB	900GB
1	SAS (7.2krpm)	—	—	—	—	—	—
2	SAS (10krpm)	—	—	—	80M	—	120M
3	SAS (15krpm)	—	30M	—	—	—	—
4	Flash Drive	20M	—	45M	—	50M	—
5	Flash Module Drive	—	—	—	—	—	—

Table 3.22.1.2-1 PDEV Erase completion expectation time (2/2)

No.	Kind of PDEV	1.2TB	1.6TB	3TB	3.2TB	4TB
1	SAS (7.2krpm)	—	—	450M	—	480M
2	SAS (10krpm)	140M	—	—	—	—
3	SAS (15krpm)	—	—	—	—	—
4	Flash Drive	—	—	—	—	—
5	Flash Module Drive	—	15M	—	30M	—

Table 3.22.1.2-2 PDEV Erase TOV (1/2)

No.	Kind of PDEV	200GB	300GB	400GB	600GB	800GB	900GB
1	SAS (7.2krpm)	—	—	—	—	—	—
2	SAS (10krpm)	—	—	—	200M	—	300M
3	SAS (15krpm)	—	100M	—	—	—	—
4	Flash Drive	70M	—	120M	—	130M	—
5	Flash Module Drive	—	—	—	—	—	—

Table 3.22.1.2-2 PDEV Erase TOV (2/2)

No.	Kind of PDEV	1.2TB	1.6TB	3TB	3.2TB	4TB
1	SAS (7.2krpm)	—	—	1000M	—	1230M
2	SAS (10krpm)	320M	—	—	—	—
3	SAS (15krpm)	—	—	—	—	—
4	Flash Drive	—	—	—	—	—
5	Flash Module Drive	—	60M	—	80M	—

Copyright © 2012, Hitachi, Ltd.

3.22.1.3 Influence in combination with other maintenance operation

The influence on the maintenance operation during executing PDEV Erase becomes as follows.

Table 3.22.1.3-1 PDEV Replace

No.	Object part	Influence	Countermeasure
1	Replace from SVP as for PDEV that does PDEV Erase.	PDEV Erase terminates abnormally.	—
2	Replace from SVP as for PDEV that does not PDEV Erase.	Nothing	—
3	User Replace	Please do not execute the user replacement during PDEV Erase.	Please execute it after completing PDEV Erase.

Table 3.22.1.3-2 DKB Replace

No.	Object part	Influence	Countermeasure
1	DKB connected with PDEV that is executed PDEV Erase	[SVP4198W] may be displayed. The DKB replacement might fail by [ONL2412E] when the password is entered. (*2)	<SIM4c2xxx/4c3xxx about this PDEV is not reported> Please replace PDEV (to which Erase is done) to new service parts. (*1) The DKB replacement might fail by [ONL2412E] when the password is entered. (*2)
2	DKB other than the above	Nothing	Nothing

Table 3.22.1.3-3 MPB Replace/MPB Removal

No.	Object part	Influence	Countermeasure
1	MPB that is executed PDEV Erase	[SVP4198W] may be displayed. The MPB replacement might fail by [ONL2412E] when the password is entered. (*2)	<SIM4c2xxx/4c3xxx about this PDEV is not reported> Please replace PDEV (to which Erase is done) to new service parts. (*1) The MPB replacement might fail by [ONL2412E] when the password is entered. (*2)
2	MPB other than the above	Nothing	Nothing

 Copyright © 2012, Hitachi, Ltd.

Table 3.22.1.3-4 ENC Replace

No.	Object part	Influence	Countermeasure
1	ENC connected with DKB connected with HDD that does PDEV Erase	[SVP4198W] may be displayed. The ENC replacement might fail by [ONL2788E] [ONL3395E] when the password is entered. (*2)	<SIM4c2xxx/4c3xxx about this PDEV is not reported> Please replace PDEV (to which Erase is done) to new service parts. (*1) The ENC replacement might fail by [ONL2788E][ONL3395E] when the password is entered. (*2)
2	ENC other than the above	Nothing	Nothing

Table 3.22.1.3-5 PDEV installation/Removal

No.	Object part	Influence	Countermeasure
1	ANY	Installation/De-installation might fail by [SVP739W].	Please wait for the Erase completion or replace PDEV (to which Erase is done) to new service parts. (*1)

Table 3.22.1.3-6 Exchanging microcode

No.	Object part	Influence	Countermeasure
1	DKC MAIN	[SVP0732W] may be displayed. Microcode exchanging might fail by [SMT2433E], when the password is entered. (*2)	Please wait for the Erase completion or replace PDEV (to which Erase is done) to new service parts. (*1)
2	HDD	[SVP0732W] may be displayed. Microcode exchanging might fail by [SMT2433E], when the password is entered. (*2)	Please wait for the Erase completion or replace PDEV (to which Erase is done) to new service parts. (*1)

Table 3.22.1.3-7 LDEV Format

No.	Object part	Influence	Countermeasure
1	ANY	There is a possibility that PATH-Inline fails. There is a possibility that the cable connection cannot be checked when the password is entered.	Please wait for the Erase completion or replace PDEV (to which Erase is done) to new service parts. (*1)

 Copyright © 2012, Hitachi, Ltd.

Table 3.22.1.3-8 PATH-Inline

No.	Object part	Influence	Countermeasure
1	DKB connected with PDEV that is executed PDEV Erase	There is a possibility of detecting the trouble by PATH-Inline.	Please wait for the Erase completion or replace PDEV (to which Erase is done) to new service parts. (*1)

Table 3.22.1.3-9 PS/OFF

No.	Object part	Influence	Countermeasure
1	ANY	PDEV Erase terminates abnormally.	<SIM4c2xxx/4c3xxx about this PDEV is not reported> Please wait for the Erase completion or replace PDEV (to which Erase is done) to new service parts. (*1)

*1: When PDEV that stops PDEV Erase is installed into DKC again, it might fail by Spin-up failure.

*2: It is not likely to be able to maintain it when failing because of concerned MSG until PDEV Erase is completed or terminates abnormally.

 Copyright © 2012, Hitachi, Ltd.

3.22.1.4 Notes of various failures

Notes of the failure during PDEV Erase become as follows.

No.	Failure	Object part	Notice	Countermeasure
1	B/K OFF/ Black Out	Drive BOX (DB)	There is a possibility that PDEV Erase fails due to the failure.	Please replace PDEV of the Erase object to new service parts after P/S on.
2		DKC	Because watch JOB of Erase disappears, it is not possible to report on normality/abnormal termination SIM of Erase.	Please replace PDEV of the Erase object to new service parts after P/S on.
3	MP failure	MPB	[E/C 9470 is reported at the MP failure] JOB of the Erase watch is reported on E/C 9470 when Abort is done due to the MP failure, and ends processing. In this case, it is not possible to report on normality/abnormal termination SIM of Erase.	Please replace PDEV of the Erase object to new service parts after the recovery of MP failure.
4			[E/C 9470 is not reported at the MP failure] It becomes impossible to communicate with the controller who is doing Erase due to the MP failure. In this case, it becomes TOV of watch JOB with E/C 9450, and reports abnormal SIM.	Please replace PDEV to new service parts after judging the Erase success or failure after it waits while TOV of PDEV Erase after the recovery of MP.

3.23 Open platform

3.23.1 GENERAL

3.23.1.1 Product Outline and Features

The open platform optional feature can assign a partial or full of disk volume area of the DKC for the Open system hosts by installing Channel Blade (CHB) to the disk controller (hereinafter called DKC). This function enables a use of high reliable and high performance storage system realized by the DKC for an open platform or Fibre system environment. This also provides the customers with a flexible and optimized system construction capability for their system expansion and migration. In Open system environment, Fibre Channel (FC) can be used as a channel interface.

3.23.1.1.1 Fibre attachment option (FC)

Some of the major features of this Fibre attachment option are listed below.

(1) Fibre interface connectivity
Fibre channel interface can be mounted as one controller. This enables multiplatform system users to share the high reliable and high performance resource realized by the DKC.

The SCSI interface is complied with ANSI SCSI-3, a standard interface for various peripheral devices for open systems. Thus, the DKC can be easily connected to various open-market Fibre host systems (e.g. Workstation servers and PC servers).

DW700 can be connected to open system via Fibre interface by installing Fibre Blade (DF-F850-HF8GR).
Fibre connectivity is provided as channel option of DW700.
Fibre Blade can be installed in any CHB location of DW700.

(2) Fast and concurrent data transmission
Data can be read and written at a maximum speed of 8 Gbps with use of Fibre interface.
All of the Fibre ports can transfer data concurrently too.

(3) All Fibre configuration (only for FC)
All Fibre configuration is also allowed either with one CHB pair or two, three or full of 12 CHB pairs configuration.
These will provide more flexible use of the storage system for open system environment.

(4) High performance
The DKC has two independent areas of nonvolatile cache memory and this mechanism also applies to the Fibre attachment option. Thus, compared with a conventional disk array controller used for open systems and not having a cache, this storage system has the following outstanding characteristics:

① Cache data management by LRU control
② Adoption of DFW (DASD Fast Write)
③ Write data duplexing
④ Nonvolatile cache

(5) High availability
The DKC is fault-tolerant against even single point of failure in its components and can successively read and write data without stopping the system. Fault-tolerance against path failures depends on the multi-path configuration support of the host system too.

(6) High data reliability
The Fibre attachment option automatically creates a guarantee code of a unique eight byte data, adds it to host data, and writes it onto the disk as data. The data guarantee code is checked automatically on the internal data bus of the DKC to prevent data errors due to array-specific data distribution or integration control. Thus, the reliability of the data improves.

Copyright © 2012, 2013, Hitachi, Ltd.

(7) TrueCopy Support (only for FC)
TrueCopy is a function to realize the duplication of open system data by connecting the two DW700 storage systems or inside parts of a single DW700 using the Fibre.
This function enables the construction of a backup system against disasters by means of the duplication of data including those of the host system or the two volumes containing identical data to be used for different purposes. (As FCoE does not support the initiator port, TrueCopy configuration using FCoE cannot be created.)

(8) HAM (High Availability Manager) (only for FC)
HAM is a function that provides a storage system that can instantly continue the user operation in an unplanned termination of the DKC due to external causes such as power failure in a place where the storage device is placed. HAM is formed in combination with the alternate path software (HDLM) that supports this function, and prevents the storage device termination due to any cause beyond control that affects the user operation, and increases the availability of the entire information system including the operation application.

Copyright © 2012, Hitachi, Ltd.

3.23.1.2 Basic Specifications

The basic specifications of the Fibre attachment are shown in Table 3.23.1.2-1.

Table 3.23.1.2-1 Basic specifications

		Specification
Item		FC
Host Channel	Max. # of Channels	48 (*1)
	Max. # of concurrent paths	256 (*1)
Data transfer		DF-F850-HF8GR: 2, 4, 8 Gbps
RAID level		RAID6/RAID5/RAID1
RAID configuration		RAID6
		RAID5
		RAID1
Cache capacity	minimum	32 G byte/System
	maximum	256 G byte/System
	additional unit	32 G byte

*1: In case of all CHBs configration.

 Copyright © 2012, Hitachi, Ltd.

3.23.1.3 Terminology

(1) Arbitrated Loop
A configuration that allows multiple ports to be connected serially.

(2) CHB
CHannel Blade. A hardware package to connect with a channel interface.

(3) Command descriptor block (CDB)
A command block in SCSI interface used to send requests from the initiator to a target.

(4) DKB
DisK Blade. A hardware package which controls disk drives within a DKC.

(5) DKC
DisK Controller. A disk controller unit consisting of CHB, DKB, Cache and other components except DB.

(6) DB
Disk Box. Disk drives units.

(7) Fabric
The entity which interconnects various N-Ports attached to it and is capable of routing frames.

(8) FAL
File Access Library: A program package and provided as a program product for Cross-OS File Exchange.

(9) FCU
File Conversion Utility: A program package and provided together with FAL for Cross-OS File Exchange.

(10) HMBR
Hitachi Multiplatform Backup/Restore.

(11) HA configuration
High Availability configuration

Copyright © 2012, Hitachi, Ltd.

(12) Initiator
The OPEN device (usually, a host computer) that requests another OPEN device to operate.

(13) Logical unit (LU)
The logical unit of division of the storage system data area accessible from SCSI interface.

(14) Logical unit number (LUN)
Identifier for a logical unit. LUN0-2048 can be assigned.

(15) Logical volume or logical device (LDEV)
The disk pack image, formed on an array disk.

(16) Point-to-Point
A configuration that allows two ports to be connected serially.

(17) Open device
Collectively refers to the host computer, peripheral control units, and intelligent peripherals that are connected to Fibre channel.

(18) Target
An Open device (usually, the DKC) that operates at the request of the initiator.

(19) VENDOR UNIQUE or VU
A manufacturer- or device-unique definable bit, byte, field, or code value.

(20) Initiator Port
A port-type used for MCU port of TrueCopy function.

(21) RCU Target Port
A port-type used for RCU port of TrueCopy function.
This port allows LOGIN of host computers and MCUs.

(22) Target port
A port-type which is different from “Initiator Port” and “RCU Target Port”.
This port is a normal target port which is used without configuration of TrueCopy.
This “Target port” allows LOGIN of host computers. It does not allow LOGIN of MCUs.

(23) External Port
Port attribute set when using it as initiator of Universal Volume Manager function.

Copyright © 2012, Hitachi, Ltd.

3.23.1.4 Notice about maintenance operations

There are some notices about Fibre maintenance operations.

(1) Before LUN path configuration is changed, Fibre I/O on the related Fibre port must be stopped.

(2) Before Fibre channel Blade or LDEV is de-installed, the related LUN path must be de-installed.

(3) Before Fibre channel Blade is replaced, the related Fibre I/O must be stopped.

(4) When Fibre-Topology information is changed, pull out a Fibre cable between the port and SWITCH and put it back again. Before a change of Fibre-Topology information, pull out Fibre cable and put it back after completing the change.

 Copyright © 2012, Hitachi, Ltd.

3.23.2 Interface Specification

3.23.2.1 Fibre Physical Interface Specification

The physical interface specification supported for Fibre Channel (FC) is shown in Table 3.23.2.1-1 to 3.23.2.1-3.

Table 3.23.2.1-1 Fibre Physical specification

No.	Item		Specification	Remark
1	Host interface	Physical interface	Fibre Channel	FC-PH,FC-AL
		Logical interface	SCSI-3	FCP,FC-PLDA
			Fibre	FC-AL
2	Data Transfer Rate	Optic Fibre cable	2, 4, 8 Gbps	
3	Cable Length	Optic single mode Fibre	10km	Longwave laser
		Optic multi mode Fibre	150 m/300 m/500 m	Shortwave laser
4	Connector Type		LC	—
5	Topology		NL-Port (FC-AL) F-Port FL-Port	—
6	Service class		3	—
7	Protocol		FCP	—
8	Transfer code		8B/10B translate	—
9	Number of hosts		255/Path	—
10	Number of host groups		255/Path	—
11	Maximum number of LUs		2048/Path	—
12	PORT/PCB		4 Port	—

SP Mode : Standard Performance Mode
HP Mode : High Performance Mode

 Copyright © 2012, Hitachi, Ltd.

Table 3.23.2.1-2 Port name

Cluster	PCB Name	PCB Location	Port Name	Cluster	PCB Name	PCB Location	Port Name
CLS1	CHB-1A (BASIC)	1A	1A	CLS2	CHB-2A (BASIC)	2A	2A
			3A				4A
			5A				6A
			7A				8A
	CHB-1B (ADD1)	1B	1B		CHB-2B (ADD1)	2B	2B
			3B				4B
			5B				6B
			7B				8B
	CHB-1C (ADD2)	1C	1C		CHB-2C (ADD2)	2C	2C
			3C				4C
			5C				6C
			7C				8C
	CHB-1D (ADD3)	1D	1D		CHB-2D (ADD3)	2D	2D
			3D				4D
			5D				6D
			7D				8D
	CHB-1E (DKB Slot)	1E	1E		CHB-2E (DKB Slot)	2E	2E
			3E				4E
			5E				6E
			7E				8E
	CHB-1F (DKB Slot)	1F	1F		CHB-2F (DKB Slot)	2F	2F
			3F				4F
			5F				6F
			7F				8F

3.23.3 CONFIGURATION

3.23.3.1 System Configurations

3.23.3.1.1 All Fibre Configuration

The DKC can also have the All Fibre interface configuration installed only by CHB adapters. The possible system configurations for the All Fibre configuration are shown below. (only for FC. In case of FCoE, this configuration is not supported.)

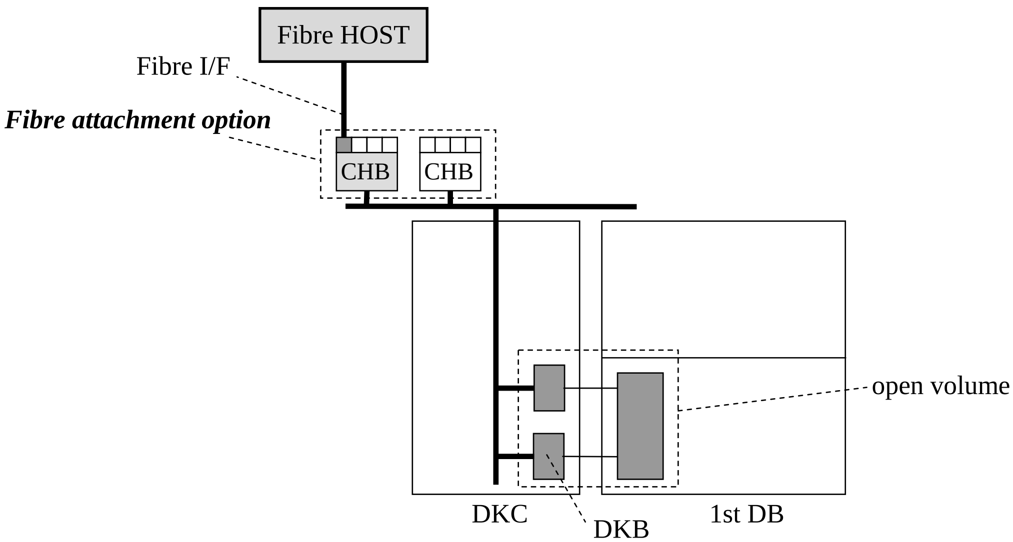


Fig. 3.23.3.1.1-1 Minimum system configuration for All Fibre

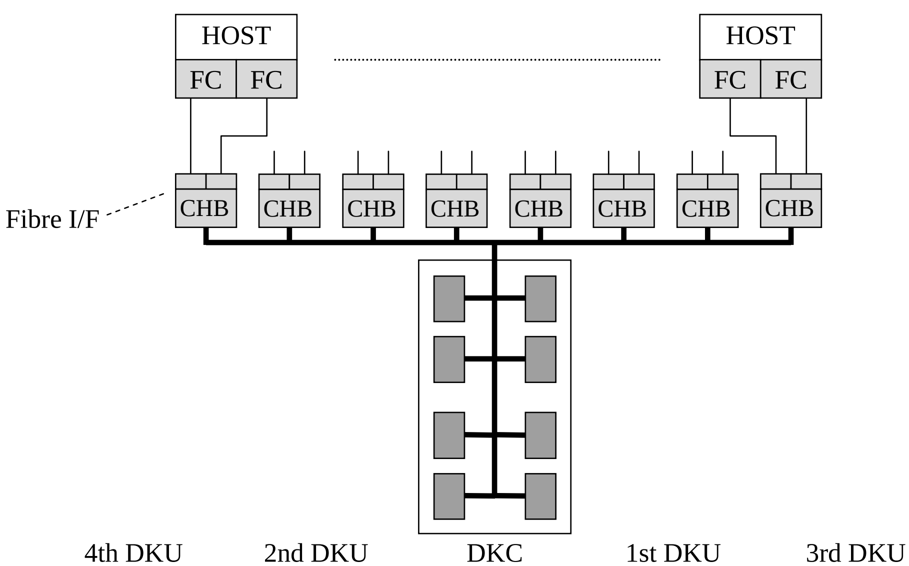


Fig. 3.23.3.1.1-2 Maximum system configuration example for All Fibre

Copyright © 2012, Hitachi, Ltd.

3.23.3.2 Channel Configuration

3.23.3.2.1 Fibre Channel Configuration

The Fibre channel Blade (CHB) PCBs must be used in sets of two. Up to 12 PCBs of the CHB can be installed in the DKC in total.

Copyright © 2012, Hitachi, Ltd.

3.23.3.3 Fibre Addressing

Each Fibre device can set a unique Port-ID number within the range from 1 to EF.
An addressing from the Fibre host to the Fibre volume in the DKC can be uniquely defined with a nexus between them. The nexus through the Initiator (host) ID, the Target (CHB port) ID, and LUN (Logical Unit Number) defines the addressing and access path. The maximum number of LUNs that can be assigned to one port is 2048.

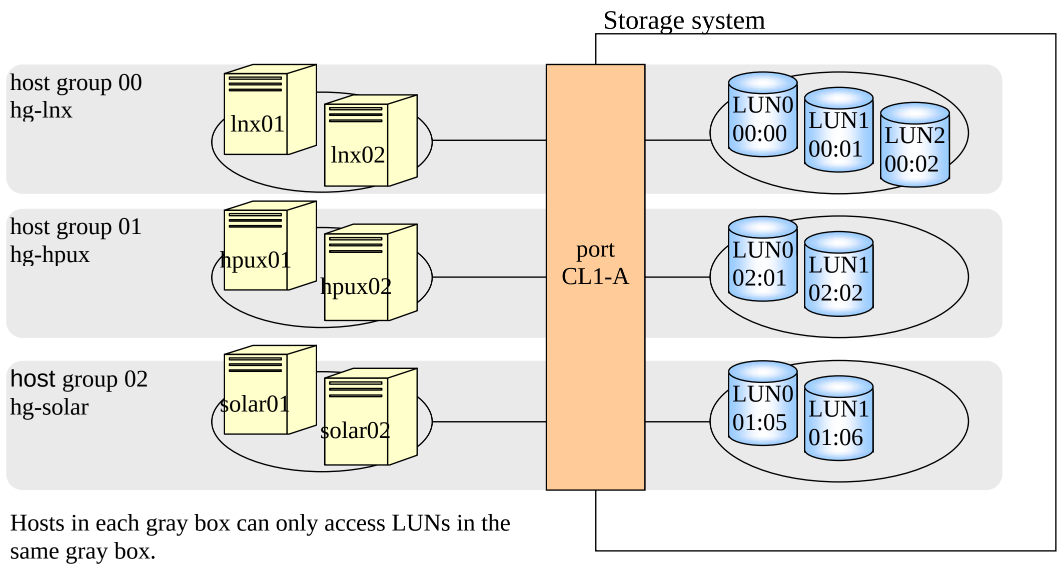
The addressing configuration is shown in the Fig. 3.23.3.3.3-1.

3.23.3.3.1 Number of Hosts

For Fibre channel, the number of connectable hosts is limited to 256 per Fibre port. (FC)
For MCU port of TrueCopy function, this limitation is as follows:
The number of MCU connections is limited to 16 per RCU Target port. (only for FC)

3.23.3.3.2 Number of Host Groups

You can define a host group admitted access for the some LU by LUN Security as a Host Group. For example, the two hosts in the hg-lnx group can only access the three LUs (00:00, 00:01, and 00:02). The two hosts in the hg-hpux group can only access the two LUs (02:01 and 02:02). The two hosts in the hg-solar group can only access the two LUs (01:05 and 01:06).



 Copyright © 2012, Hitachi, Ltd.

3.23.3.3.3 LUN (Logical Unit Number)

LUNs can be assigned from 0 to 2047 to each Fibre Port.

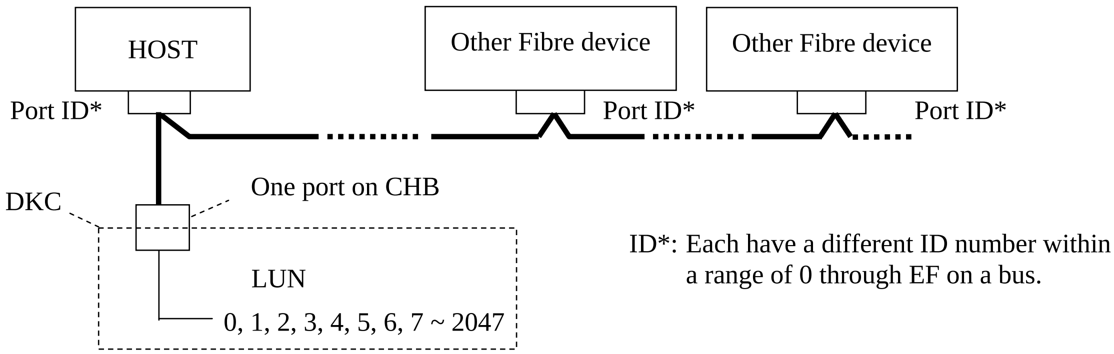


Fig. 3.23.3.3.3-1 Fibre addressing configuration from Host

3.23.3.3.4 PORT INFORMATION

A PORT address and the Topology can be set as PORT INFORMATION. The value of PORT address is EF and can be changed by user. Topology information is selected from “Fabric”, “FC-AL” or “Point to point”.

 Copyright © 2012, Hitachi, Ltd.

3.23.3.4 Logical Unit

3.23.3.4.1 Logical Unit Specification

The specifications of Logical Units supported and accessible from Open system hosts are defined in the Table 3.23.3.4.1-1.

Table 3.23.3.4.1-1 LU specification

No	Item		Specification	
1	Access right		Read/Write	
2	Logical Unit (LU) size	G byte (10^9)	OPEN-V×n	
		G byte ($1{,}024^3$)	—	
3	Block size		512 Bytes	
4	# of blocks		—	
5	LDEV emulation name		OPEN-V	

*1: “0” is added to the emulation type of the V-VOLs (e.g. OPEN-0V).
When you create a Thin Image pair, specify the volume whose emulation type is displayed with “0” like OPEN-0V as the S-VOL.

Copyright © 2012, Hitachi, Ltd.

3.23.3.4.2 Logical Unit Mapping of Fibre

Each volume name, such as OPEN-V is also used as an emulation type name to be specified for each ECC group. When the emulation type is defined on an ECC group, Logical volumes (LDEVs) are automatically allocated to the ECC group from the specified LDEV#. After creating LDEVs, each LUN of Fibre port will be mapped on any location of LDEV within DKC. This setting is performed by SVP operation or Web Console operation (option).

This flexible LU and LDEV mapping scheme enables the same logical volume to be set to multiple paths so that the host system can configure a shared volume configuration such as a High Availability (HA) configuration. In the shared volume environment, however, some lock mechanism need to be provided by the host systems.

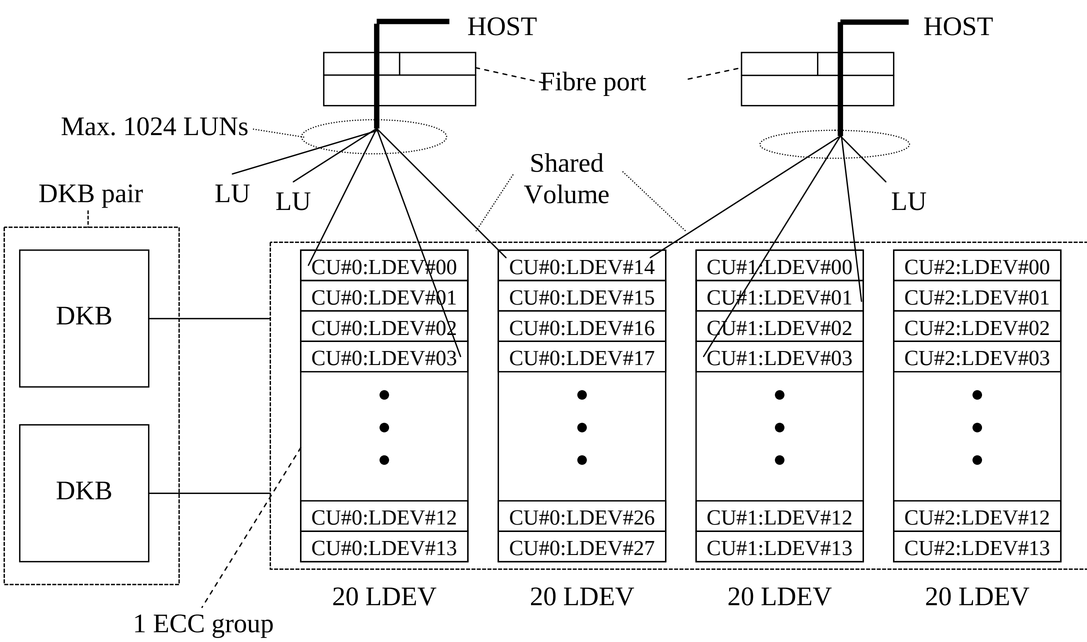


Fig. 3.23.3.4.2-1 LDEV and LU mapping for Fibre volume

Copyright © 2012, Hitachi, Ltd.

3.23.3.4.3 Logical Unit Size Expansion (LUSE) Function

(1) Outline

This is a function to show the host the continuous LDEV of a volume exclusive for open system as a virtually large LU.
In the former configuration, one LU is one LDEV, but this expanding function can enlarge the LU size up to 265.8 G byte for example by showing the host two or more continuous LDEVs as a single LU.
Many LUs have been needed to cover the entire capacity of a storage system before, but this function enables a small number of LUs to cover it from the viewpoint of host interface.
The MCU port (Initiator port) of TrueCopy function does not support LU size expansion.

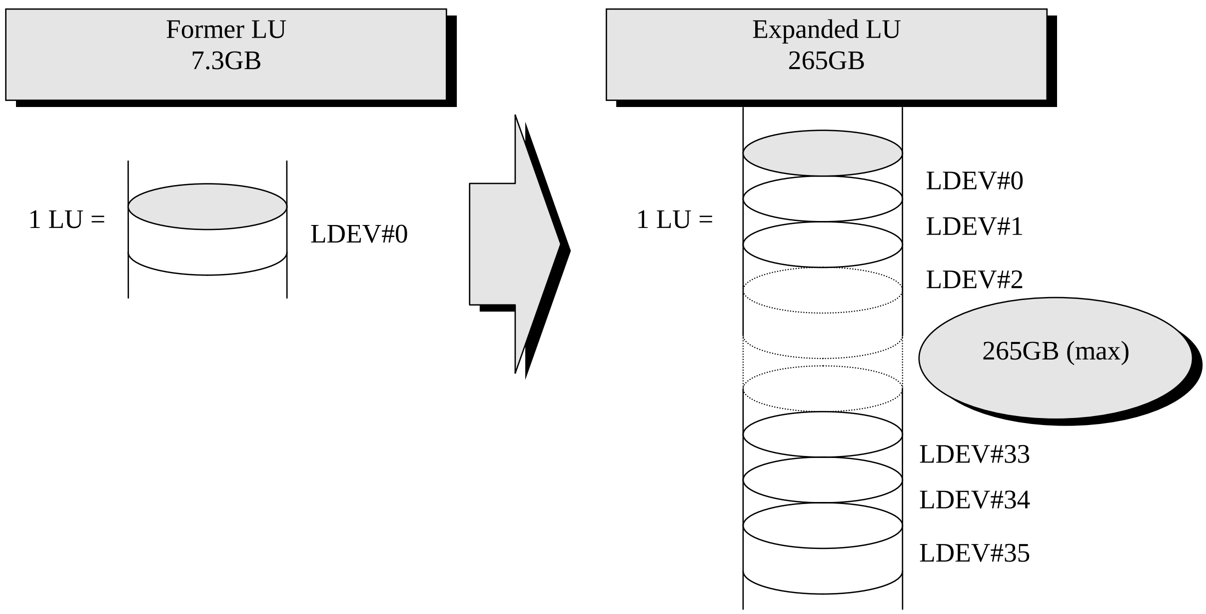


Fig. 3.23.3.4.3-1 Example of LU Size Expansion Function

Copyright © 2012, Hitachi, Ltd.

(2) Specifications

Table 3.23.3.4.3-1 shows specifications for the LUSE. (1 KB = 1024 Byte)

Table 3.23.3.4.3-1 LUSE Specification

LDEV Capacity	46.0 M byte ~ 2.8 T byte
Number of connectable LDEVs/LU	2 to 36
LU Capacity	92.0 M byte to 60 T byte
Product name for responding to INQUIRY	OPEN-V×n

n: Number of connected LDEVs

*1: LDEVs can be connected with different CU number and capacity.

Copyright © 2012, Hitachi, Ltd.

(3) Effects and restrictions of LUSE function

(a) Effects

- Restrictions of usable capacity owing to the number of the usable hosts is released.
  -Restriction of the host capacity
  -Restriction of capacity owing to restriction of the number of LUs of the HA software
- The disk connection function on the host side such as VxVM becomes unnecessary.
- Effect of LU size extending with LDEV.
  -LU of optional size can be configured.
  -The load of PDEV can be dispersed by the LUSE configuration of CV dispersed in ECC.
  -Performance can be improved by increasing the multiplex frequency of LDEV.

(b) Restrictions

- Some OSs are slow in disk accesses handling large data and may not be usable depending on environment. (Example: AIX is slow in accesses handling data larger than 2GB.)
- The capacity should be determined as necessary in a system designed to achieve a high-speed operation by making the LUs perform multiple operation.

(4) Notes on use

When the LU is expanded, the following restrictions are added to a case where no expansion is made, such as a change in capacity seen from the open host owing to the specification of the expansion.

(a) The LU size cannot be changed while the LU is being used by the host. If you want to change the LU size, the host must be rebooted once. If the LU size once set is to be changed, shut down the host, change the LU size, then start up the host again.

(b) If an LU to being used or expanded is reconfigured in a new configuration or as a expanded LU, data which had been used will be lost.
Perform physical replacement work of the disk including data backup, separation of the former LU, LU connection after the configuration change, and restoration of backup data.

(c) When an LDEV in the LU is blocked, an LU blocking error does not occur unless an access is made to the blocked LDEV. When the access to the blocked LDEV is made, a blocking error occurs in the expanded LU.

(d) The maintenance procedure when an error such as an LDEV blocking occurs is the same as before. Check the LDEV status from the SVP and perform the maintenance considering the relation between the LDEV and the LU.

Copyright © 2012, Hitachi, Ltd.

3.23.3.4.4 LUN Security

(1) Outline
This function connects various kinds of servers into a segregated, secure environment via the switch in the Fibre channel port, and thus enables the storage and the server to be used in the SAN environment.
The MCU (initiator) port of TrueCopy does not support this function.

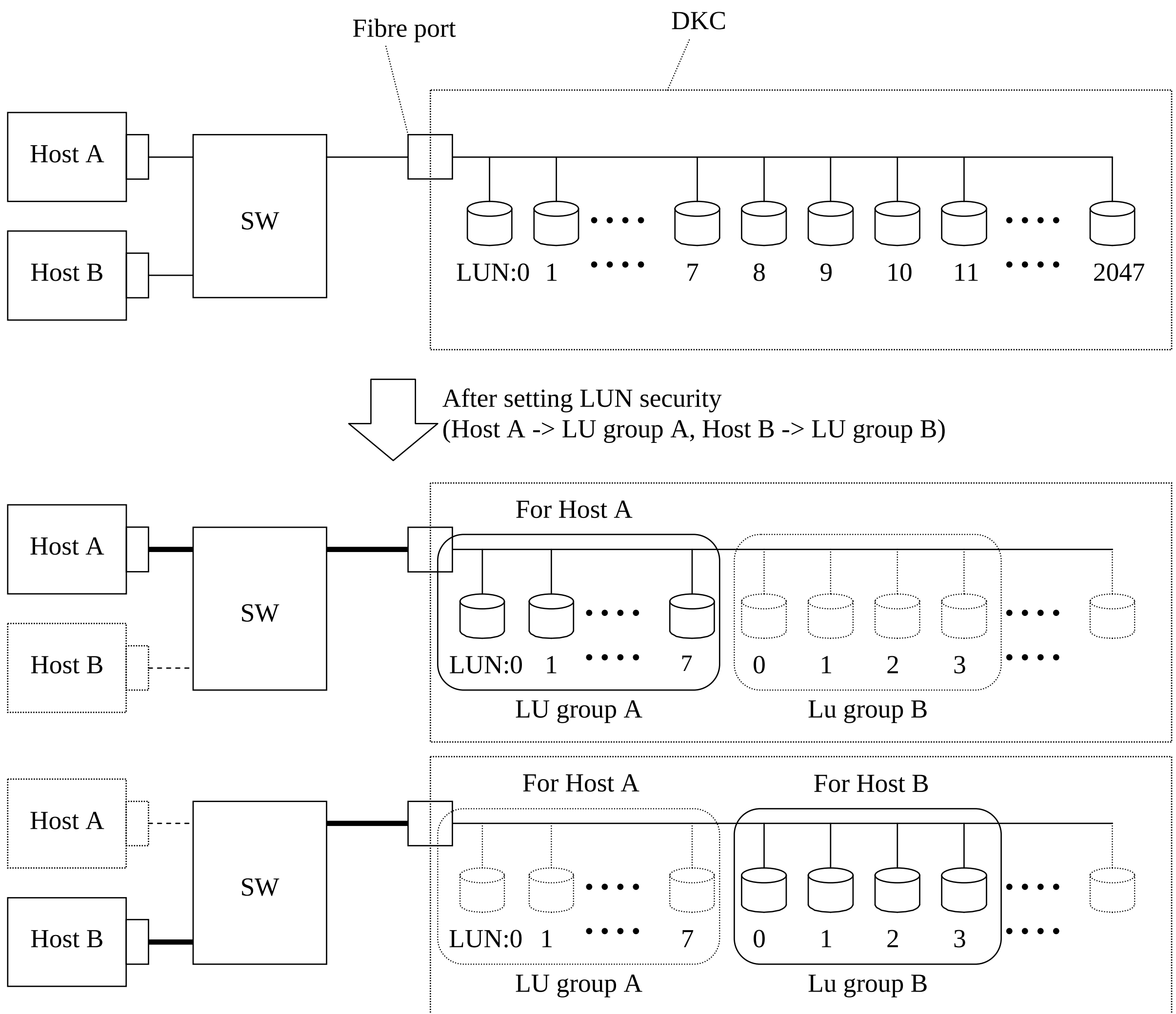


Fig. 3.23.3.4.4-1 LUN Security

 Copyright © 2012, 2013, Hitachi, Ltd.

3.23.3.5 Volume Specification

3.23.3.5.1 Volume Specification

Model Number of Disk drive and supported RAID Level are shown in Table 3.23.3.5.1-1.

Table 3.23.3.5.1-1 List of DW700 Model number

Model Number	Disk drive model	RAID Level
DF-F850-3HGSSH	DKS5C-K300SS	RAID1 (2D+2D) RAID5 (3D+1P, 7D+1P) RAID6 (6D+2P, 14D+2P)
DF-F850-6HGSS	DKR5C-J600SS	
	DKR5D-J600SS	
	DKS5D-J600SS	
	DKS5E-J600SS	
DF-F850-9HGSS	DKR5D-J900SS	
	DKS5D-J900SS	
	DKS5E-J900SS	
DF-F850-12HGSS	DKR5E-J1R2SS	
DF-F850-2HGDM	SLB5A-M200SS	
	SLR5B-M200SS	
DF-F850-4HGDM	SLB5A-M400SS	
	SLR5B-M400SS	
DF-F850-8HGDM	SLB5A-M800SS	
DKC-F710I-1R6FM	NFH1A-P1R6SS	
DF-F850-3TNX	DKR2D-H3R0SS	
	DKR2E-H3R0SS	
	DKS2D-H3R0SS	
	DKS2E-H3R0SS	
DKC-F710I-3R2FM	NFH1B-P3R2SS	
DF-F850-4TNX	DKR2E-H4R0SS	
	DKS2E-H4R0SS	

NOTE: As for RAID1, the concatenation of two parity groups is possible (8HDDs).
In this case the number of volumes required is doubled.
Two concatenation and four concatenation (16HDDs and 32HDDs) of the RAID Groups are possible for RAID5 (7D+1P).
In this case, the number of volumes becomes twice or four times.
When OPEN-V is set in the parity group of the above-mentioned connection configuration, the maximum volume size becomes the parity cycle size of the source (2D+2D) or (7D+1P). It does not become twice or four times.

NOTE: The storage system capacity is different from one of SVP, because of 1GB=1000^3byte calculation.

 Copyright © 2012, 2013, Hitachi, Ltd.

Table 3.23.3.5.1-2 The relation between capacity, RAID level and DB type

Drive type		Capacity		Number of LDEV
		MB	Logical BlocKSR	
RAID5 (3D+1P)	DF-F850-3HGSSH	824536.5	1688650752	1
	DF-F850-6HGSS	1649074.5	3377304576	1
	DF-F850-9HGSS	2473764.0	5066268672	1
	DF-F850-12HGSS	3298149.0	6754609152	2
	DF-F850-2HGDM	563400.0	1153843200	1
	DF-F850-4HGDM	1126801.5	2307689472	1
	DF-F850-8HGDM	2253604.5	4615382016	1
	DKC-F710I-1R6FM	5033157.0	10307905536	2
	DF-F850-3TNX	8403129.0	17209608192	3
	DKC-F710I-3R2FM	10066320.0	20615823360	4
	DF-F850-4TNX	11204172.0	22946144256	4
RAID5 (7D+1P)	DF-F850-3HGSSH	1923918.5	3940185088	1
	DF-F850-6HGSS	3847837.0	7880370176	2
	DF-F850-9HGSS	5772116.0	11821293568	2
	DF-F850-12HGSS	7695681.0	15760754688	3
	DF-F850-2HGDM	1314600.0	2692300800	1
	DF-F850-4HGDM	2629203.5	5384608768	1
	DF-F850-8HGDM	5258407.0	10769217536	2
	DKC-F710I-1R6FM	11744026.0	24051765248	4
	DF-F850-3TNX	19607301.0	40155752448	7
	DKC-F710I-3R2FM	23488080.0	48103587840	8
	DF-F850-4TNX	26143078.5	53541024768	9
RAID1 (2D+2D)	DF-F850-3HGSSH	549691.0	1125767168	1
	DF-F850-6HGSS	1099383.0	2251536384	1
	DF-F850-9HGSS	1649176.0	3377512448	1
	DF-F850-12HGSS	2198766.0	4503072768	1
	DF-F850-2HGDM	375600.0	769228800	1
	DF-F850-4HGDM	751201.0	1538459648	1
	DF-F850-8HGDM	1502403.0	3076921344	1
	DKC-F710I-1R6FM	3355438.0	6871937024	2
	DF-F850-3TNX	5602088.0	11473076224	2
	DKC-F710I-3R2FM	6710883.0	13743888384	3
	DF-F850-4TNX	7469451.0	15297435648	3

(To be continued)

Copyright © 2012, 2013, Hitachi, Ltd.

(Continued from the preceding page)

Drive type		Capacity		Number of LDEV
		MB	Logical BlocKSR	
RAID6 (6D+2P)	DF-F850-3HGSSH	1649073.0	3377301504	1
	DF-F850-6HGSS	3298146.0	6754603008	2
	DF-F850-9HGSS	4947528.0	10132537344	2
	DF-F850-12HGSS	6596298.0	13509218304	3
	DF-F850-2HGDM	1126800.0	2307686400	1
	DF-F850-4HGDM	2253603.0	4615378944	1
	DF-F850-8HGDM	4507206.0	9230757888	2
	DKC-F710I-1R6FM	10066308.0	20615798784	4
	DF-F850-3TNX	16806258.0	34419216384	6
	DKC-F710I-3R2FM	20132637.0	41231640576	7
	DF-F850-4TNX	22408344.0	45892288512	8
RAID6 (14D+2P)	DF-F850-3HGSSH	3847840.5	7880377344	2
	DF-F850-6HGSS	7695681.0	15760754688	3
	DF-F850-9HGSS	11544232.0	23642587136	4
	DF-F850-12HGSS	15391355.0	31521495040	5
	DF-F850-2HGDM	2629203.5	5384608768	1
	DF-F850-4HGDM	5258410.5	10769224704	2
	DF-F850-8HGDM	10516800.0	21538406400	4
	DKC-F710I-1R6FM	23488024.0	48103473152	8
	DF-F850-3TNX	39214616.0	80311533568	13
	DKC-F710I-3R2FM	46976160.0	96207175680	15
	DF-F850-4TNX	52286101.0	107081934848	17

NOTE: The relationship between Capacity [MB] and Logical BlocKSR is as follows.
A[MB] = B * 512 / 1000 / 1000 (1MB = 1000^2 byte)
or
A[MB] = B * 512 / 1024 / 1024 (1MB = 1024^2 byte)
where C is the capacity by Logical BlocKSR.

Copyright © 2012, Hitachi, Ltd.

Blank Sheet

Copyright © 2012, Hitachi, Ltd.

3.23.3.6 Volume Setting

3.23.3.6.1 Setting of volume space

The procedure of volume setting is performed either by using the SVP or Storage Navigator function (optional feature).

3.23.3.6.2 LUN setting

- *LUN setting:*

  - Select the CHB, Fibre port and the LUN, and select the CU# and LDEV# to be assigned to the LUN.
  - Repeat the above procedure as needed.
    The MCU port (Initiator port) of TrueCopy function does not support this setting.

*1: It is possible to refer to the contents which is already set on the SVP display.

*2: The above setting can be done during on-line.

*3: Duplicated access paths' setting from the different hosts to the same LDEV is allowed. This will provide a means to share the same volume among host computers. It is, however, the host responsibility to manage an exclusive control on the shared volume.

Refer to the INSTALLATION SECTION for more detailed procedures.

Copyright © 2012, Hitachi, Ltd.

3.23.3.7 Host mode setting

It is necessary to set Host Mode by using SVP if you want to change a host system.
The meanings of each mode are follows.

```
*******HDS RAID Controller Models**************************
MODE 00   : Standard mode (Linux)
MODE 01   : VMWare host mode
MODE 03   : HP-UX host mode
MODE 05   : OpenVMS host mode
MODE 07   : Tru64 host mode
MODE 09   : Solaris host mode
MODE 0A   : NetWare host mode
MODE 0C   : Windows2000/2003
MODE 0F   : AIX host mode
MODE 21   : VMWare host mode (Online LUSE)
MODE 2C   : Windows2000/2003 host mode (Online LUSE)
others    : Reserved
*******************************************************
```

```
*******HP RAID Controller Models***************************
MODE 00   : Standard mode (Linux)
MODE 01   : VMWare host mode
MODE 05   : OpenVMS host mode
MODE 07   : Tru64 host mode
MODE 08   : HP-UX host mode
MODE 09   : Solaris host mode
MODE 0A   : NetWare host mode
MODE 0C   : Windows2000/2003, NonStop OS
MODE 0F   : AIX host mode
MODE 21   : VMWare host mode (Online LUSE)
MODE 2C   : Windows2000/2003 host mode (Online LUSE), NonStop OS (Online LUSE)
others    : Reserved
*******************************************************
```

Please set the HOST MODE OPTION if required.
Please see “LUN Management” (INST05-1010 to INST05-1540). Also see the operational manual for more detailed information about the alternate link and HA software.

Copyright © 2012, Hitachi, Ltd.

3.23.4 Control Function

3.23.4.1 Cache Usage

The DKC has two independent areas of non-volatile cache memory for volumes by which high reliability and high performance with the following features can be achieved.

① Cache data management by LRU control
Data that has been read out is stored into the cache and managed under LRU control. For upright transaction processing, therefore, a high cache hit ratio can be expected and a data-writing time is reduced for improved system throughput.

② Adoption of DFW (DASD Fast Write)
At the same time that the normal write command writes data into the cache, it reports the end of the write operations to a host. Data writing to disk is asynchronous with host access. The host, therefore, can execute the next process without waiting for the end of data writing to disk.

③ Write data duplexing
The same write data is stored into the two areas of a cache provided in the DKC. Thus, loss of DFW data can be avoided even one failure occurs in the cache.

④ Nonvolatile cache
Batteries and Cache Flash Memories (CFM) are installed in MAIN Blades in a DKC. Once a data has been written into a cache, even if a power interruption occurs, it always holds the data because the data is transferred to the CFM.

 Copyright © 2012, Hitachi, Ltd.

3.23.4.2 SCSI Command Multi-processing

3.23.4.2.1 Command Tag Queuing

The Command Tag Queuing function defined in the SCSI specification is supported. This function allows each Fibre port on a CHB to accept multiple SCSI commands even for the same LUN. The DKC can process those queued commands in parallel because a LUN is composed of multiple physical drives.
The MCU port (Initiator port) of TrueCopy function can not support this function because it does not support a connection with a host computer.

3.23.4.2.2 Concurrent data transfer

Fibre ports on a CHB can perform the host I/Os and data transfer with maximum 8 Gbps transfer concurrently.
This is also applied among different CHBs.
The MCU port (Initiator port) of TrueCopy function can not support this function because it does not support a connection with a host computer.

 Copyright © 2012, Hitachi, Ltd.

3.23.5 SCSI Commands

3.23.5.1 Fibre

The DASD commands defined under the SCSI-3 standards and those supported by the DKC are listed in Table 3.23.5.1-1.

Table 3.23.5.1-1 SCSI-3 DASD commands and DKC-supported commands

Group	Op Code	Name of Command	Type	○:Supported	Remarks
0	00_H	Test Unit Ready	CTL/SNS	○	
$(00_H$ -$1F_H)$	01_H	Rezero Unit	CTL/SNS	Nop	
	03_H	Request Sense	CTL/SNS	○	
	04_H	Format Unit	DIAG	Nop	
	07_H	Reassign Blocks	DIAG	Nop	
	08_H	Read (6)	RD/WR	○	
	$0A_H$	Write (6)	RD/WR	○	
	$0B_H$	Seek (6)	CTL/SNS	Nop	
	12_H	Inquiry	CTL/SNS	○	
	15_H	Mode Select (6)	CTL/SNS	○	
	16_H	Reserve	CTL/SNS	○	
	17_H	Release	CTL/SNS	○	
	18_H	Copy	–	–	
	$1A_H$	Mode Sense (6)	CTL/SNS	○	
	$1B_H$	Start/Stop Unit	CTL/SNS	Nop	
	$1C_H$	Receive Diagnostic Results	DIAG	—	
	$1D_H$	Send Diagnostic	DIAG	Nop	Supported only for self-test.
	$1E_H$	Prevent Allow Medium Removal	—	—	
	$1F_H$	Reserved code	—	—	
	Other	Vendor–unique	—	—	
1	25_H	Read Capacity (10)	CTL/SNS	○	
$(20_H$ -$3F_H)$	28_H	Read (10)	RD/WR	○	
	$2A_H$	Write (10)	RD/WR	○	
	$2B_H$	Seek (10)	CTL/SNS	Nop	
	$2E_H$	Write And Verify (10)	RD/WR	○	Supported only Write.
	$2F_H$	Verify (10)	RD/WR	Nop	
	30_H	Search Data High	—	—	
	31_H	Search Data Equal	—	—	
	32_H	Search Data Low	—	—	
	33_H	Set Limits (10)	—	—	
	34_H	Pre-Fetch (10)	—	—	
	35_H	Synchronize Cache (10)	CTL/SNS	Nop	
	36_H	Lock-Unlock Cache (10)	—	—	
	37_H	Read Defect Data (10)	DIAG	○	No defect always reported.
	38_H	Reserved code	—	—	
	39_H	Compare	—	—	

Copyright © 2012, Hitachi, Ltd.

Table 3.23.5.1-1 SCSI-2 DASD commands and DKC-supported commands (Continued)

Group	Op Code	Name of Command	Type	○:Supported	Remarks
1	$3A_H$	Copy And Verify	—	—	
(20_H -$3F_H$)	$3B_H$	Write Buffer	DIAG	○	
	$3C_H$	Read Buffer	DIAG	○	
	$3D_H$	Reserved code	—	—	
	$3E_H$	Read Long	—	—	
	$3F_H$	Write Long	—	—	
	Other	Vendor-unique	—	—	
2	40_H	Change Definition	—	—	
(40_H -$5F_H$)	41_H	Write Same	—	—	
	42_H	Unmap	CTL/SNS	○	
	$4C_H$	Log Select	—	—	
	$4D_H$	Log Sense	—	—	
	50_H	XD Write (10)	—	—	
	51_H	XP Write (10)	—	—	
	52_H	XD Read (10)	—	—	
	53_H	XD Write Read (10)	—	—	
	55_H	Mode Select (10)	CTL/SNS	○	
	56_H	Reserve (10)	CTL/SNS	○	
	57_H	Release (10)	CTL/SNS	○	
	$5A_H$	Mode Sense (10)	CTL/SNS	○	
	$5E_H$	Persistent Reserve IN	CTL/SNS	○	
	$5F_H$	Persistent Reserve OUT	CTL/SNS	○	
	Other	Reserved code	—	—	
3	$7F_H$/0001	Rebuild (32)	—	—	
(60_H -$7F_H$)	$7F_H$/0002	Regenerate (32)	—	—	
	$7F_H$/0003	XD Read (32)	—	—	
	$7F_H$/0004	XD Write (32)	—	—	
	$7F_H$/0005	XD Write Extend (32)	—	—	
	$7F_H$/0006	XD Write (32)	—	—	
	$7F_H$/0007	XD Write Read (32)	—	—	
	$7F_H$/0008	XD Write Extend (64)	—	—	
	Other	Reserved code	—	—	
4	80_H	XD Write Extend (16)	—	—	
(80_H -$9F_H$)	81_H	Rebuild (16)	—	—	
	82_H	Regenerate (16)	—	—	
	83_H	Extended Copy	CTL/SNS	○	
	84_H	Receive Copy Result	CTL/SNS	○	
	85_H	Access Control IN	—	—	
	86_H	Access Control OUT	—	—	
	88_H	Read (16)	RD/WR	○	
	89_H	Compare and Write	RD/WR	○	
	$8A_H$	Write (16)	RD/WR	○	
	$8C_H$	Read Attributes	—	—	
	$8D_H$	Write Attributes	—	—	

Table 3.23.5.1-1 SCSI-2 DASD commands and DKC-supported commands (Continued)

Group	Op Code	Name of Command	Type	○:Supported	Remarks
4 (80_H -$9F_H$)	$8E_H$	Write And Verify (16)	RD/WR	○	Supported only Write.
	$8F_H$	Verify (16)	RD/WR	Nop	
	90_H	Pre-Fetch (16)	—	—	
	91_H	Synchronized Cache (16)	CTL/SNS	Nop	
	92_H	Lock-Unlock Cache (16)	—	—	
	93_H	Write Same (16)	CTL/SNS	○	
	$9E/10_H$	Read Capacity (16)	CTL/SNS	○	
	$9E/12_H$	Get LBA Status	CTL/SNS	○	
	Other	Vendor-unique	—	—	
5 ($A0_H$ -BF_H)	$A0_H$	Report LUN	CTL/SNS	○	
	$A3_H/xx_H$	Maintenance IN	CTL/SNS	—	
	$A3_H/05_H$	Report Device Identifier	CTL/SNS	○	
	$A3_H/0A_H$	Report Target Port Groups	CTL/SNS	—	
	$A3_H/0B_H$	Report Aliases	CTL/SNS	—	
	$A3_H/0C_H$	Report Supported Operation Codes	CTL/SNS	—	
	$A3_H/0D_H$	Report Supported Task Management Functions	CTL/SNS	—	
	$A3_H/0E_H$	Report Priority	CTL/SNS	—	
	$A3_H/0F_H$	Report Timestamp	CTL/SNS	—	
	$A4_H/XX_H$	Maintenance OUT	CTL/SNS	—	
	$A4_H/06_H$	Set Device Identifier	CTL/SNS	—	
	$A4_H/0A_H$	Set Target Port Groups	CTL/SNS	○	
	$A4_H/0B_H$	Change Aliases	CTL/SNS	—	
	$A4_H/0E_H$	Set Priority	CTL/SNS	—	
	$A4_H/0F_H$	Set Timestamp	CTL/SNS	—	
	$A7_H$	Move Medium Attached	—	—	
	$A8_H$	Read (12)	RD/WR	○	
	AA_H	Write (12)	RD/WR	○	
	AE_H	Write And Verify (12)	RD/WR	○	However, only the Write operation.
	AF_H	Verify (12)	RD/WR	Nop	
	$B3_H$	Set Limits (12)	—	—	
	$B4_H$	Read Element Status Attached	—	—	
	$B7_H$	Read Defect Data (12)	CTL/SNS	○	It always reports on No defect.
	BA_H	Redundancy Group IN	—	—	
	BB_H	Redundancy Group OUT	—	—	
	BC_H	Spare IN	—	—	
	BD_H	Spare OUT	—	—	
	BE_H	Volume Set IN	—	—	
	BF_H	Volume Set OUT	—	—	
	Other	Reserved code	—	—	

Copyright © 2012, Hitachi, Ltd.

Table 3.23.5.1-1 SCSI-2 DASD commands and DKC-supported commands (Continued)

Group	Op Code	Name of Command	Type	○:Supported	Remarks
6 ($C0_H$ -DF_H)	$C0_H$~$D0_H$	Vendor-unique	—	—	
7 ($E0_H$ -FF_H)	$E8_H$	Read With Skip Mask (IBM-unique)	CTL/SNS	—	
	EA_H	Write With Skip Mask (IBM-unique)	CTL/SNS	—	
	Other	Vendor-unique	—	—	

Copyright © 2012, Hitachi, Ltd.

3.23.6 HA Software Linkage Configuration in a Cluster Server Environment

When this storage system is linked to High-Availability software (HA software) which implements dual-system operation for improved total system fault-tolerance and availability, the open system side can also achieve higher reliability on the system scale.

3.23.6.1 Example of System Configurations

(1) Hot-standby system configuration
The HA software minimizes system down time in the event of hardware or software failures and allows processing to be restarted or continued. The basic system takes a hot-standby (asymmetric) configuration, in which, as shown in the figure below, two hosts (an active host and a standby host) are connected via a monitoring communication line. In the hot-standby configuration, a complete dual system can be built by connecting the Fibre cables of the active and standby hosts to different CHB Fibre ports.

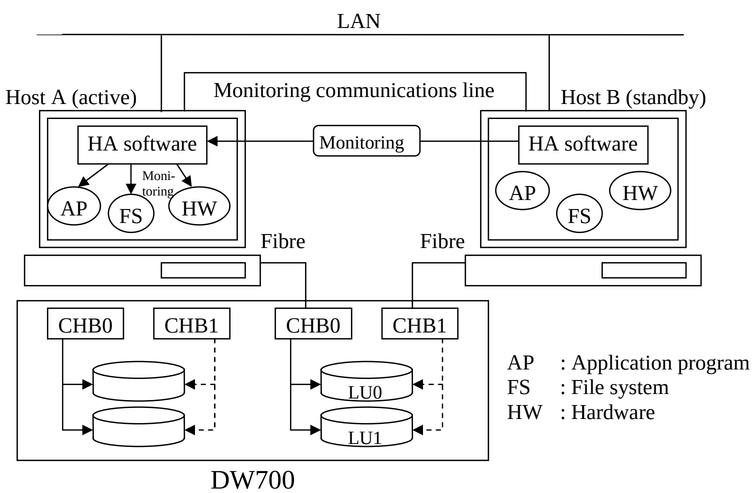


Fig. 3.23.6.1-1 Hot-standby configuration

• The HA software under the hot-standby configuration operates in the following sequence:

a. The HA software within the active host monitors the operational status of own system by using a monitoring agent and sends the results to the standby host through the monitoring communication line (this process is referred to as “heart beat transmission”). The HA software within the standby host monitors the operational status of the active host based on the received information.

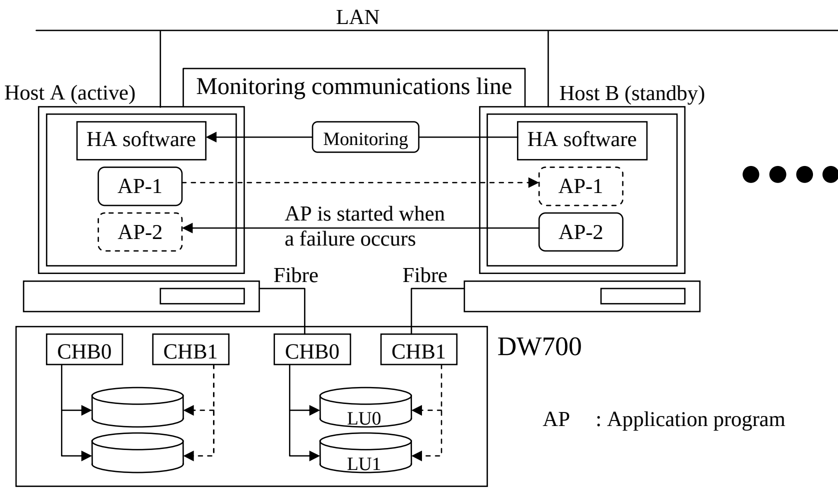
b. If an error message is received from the active host or no message is received, the HA software of the standby host judges that a failure has occurred in the active host. As a result, it transfers management of the IP addresses, disks, and other common resources, to the standby host (this process is referred to as “fail-over”).

c. The HA software starts the application program concerned within the standby host to take over the processing on behalf of the active host.

- Use of the HA software allows a processing request from a client to be taken over. In the case of some specific application programs, however, it appears to the client as if the host that was processing the task has been rebooted due to the host switching. To ensure continued processing, therefore, a login to the application program within the host or sending of the processing request may need to be executed once again.

(2) Mutual standby system configuration

In addition to the hot-standby configuration described above, a mutual standby (symmetric) configuration can be used to allow two or more hosts to monitor each other. Since this storage system has eight Fibre ports, it can, in particular, be applied to a large-scale cluster environment in which more than two hosts exist.

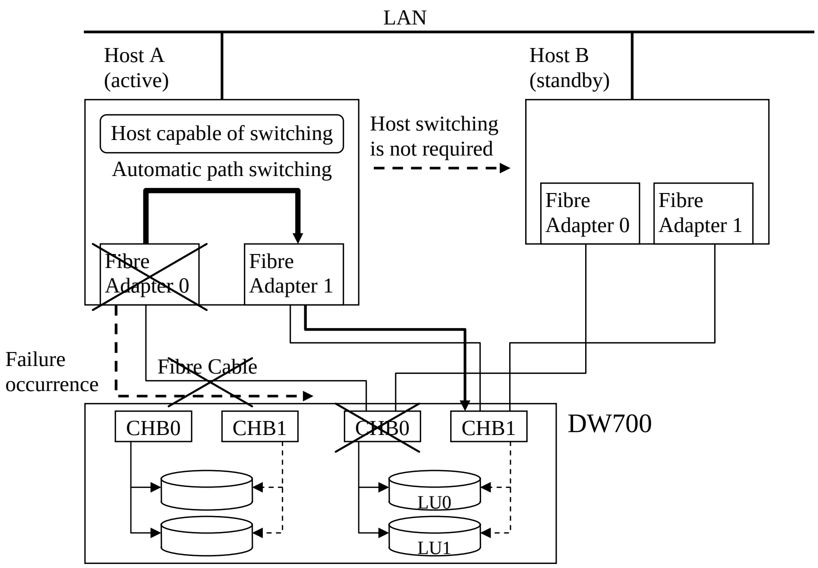


- In the mutual standby configuration, since both hosts operate as the active hosts, no resources exist that become unnecessary during normal processing. On the other hand, however, during a backup operation the disadvantages are caused that performance deteriorated and that the software configuration becomes complex.
- This storage system is scheduled to support Oracle SUN CLUSTER, Symantec Cluster server, Hewlett-Packard MC/ServiceGuard, and IBM HACMP and so on.

Copyright © 2012, Hitachi, Ltd.

3.23.6.2 Configuration Using Host Path Switching Function

When the host is interlocked with the HA software and has a path switching capability, if a failure occurs in the Fibre adapter, Fibre cable, or DKC (Fibre ports and the CHB) that is being used, automatic path switching will take place as shown below.



The path switching function enables processing to be continued without host switching in the event of a failure in the Fibre adapter, Fibre cable, array controller, or other components.

 Copyright © 2012, Hitachi, Ltd.

3.23.7 TrueCopy

3.23.7.1 Overview

The Hitachi Open Remote Copy function can remotely duplicate data (volumes) under the control of the storage system by directly connecting the two DW700s. A backup system against disasters can be constructed by installing one of the two DW700s at the main site and the other at the recovery site and configuring the HA cluster on the server side by means of the HA (High Availability) software.
This function also enables the two volumes containing identical data to be used for different purposes by duplicating data (volumes) within the same DW700 or between the two DW700s and separating the volumes in a primary-and-secondary relation at any time.
An online database can be backed up or batch programs can be executed while the database is being accessed. There are TrueCopy and Universal Replicator (UR) for TrueCopy.
The TrueCopy makes various settings and it controls operations by means of the RAID manager/TrueCopy, which runs on the open system. The RAID manager/TrueCopy provides various commands for user applications to control the TrueCopy functions. Creation of a user shell script using these commands enables the TrueCopy control being interlocked with server’s fail-over executed by the HA software.

There is Fibre channel interface of connection form between CUs.

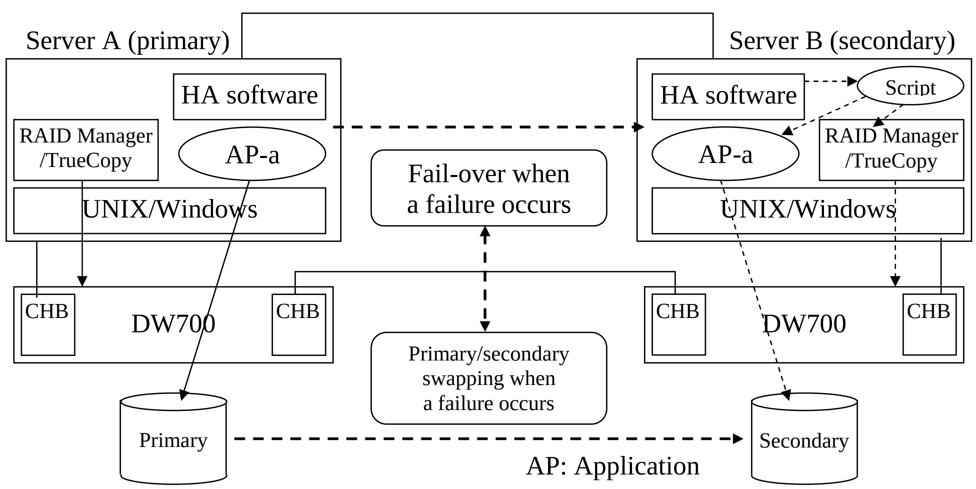


Fig. 3.23.7.1-1 Outline of TrueCopy Function and Example of Application to HA Configuration (Hot Standby Configuration)

Copyright © 2012, Hitachi, Ltd.

3.23.7.2 Basic Specifications

Basic TrueCopy specifications are shown below.

Table 3.23.7.2-1 Basic Specifications of TrueCopy

No.	Item	Description	Remarks
1	Host interface on open system side	Fibre Channel	Conforming to base platform function.
2	Supporting OS platform	Conforming to base platform function.	
3	Connection between the CUs	Fibre Channel	
4	Means for setting the paired LU	RAID Manager/TrueCopy	
		Storage Navigator	
		Web Console	
5	Number of LUs capable of the duplicated writing	Maximum 16,384 pairs	
6	LU size capable of the duplicated writing (The paired VOL must be the same DEV type.)	Normal VOL	Maximum 3TB
		DP-VOL	Maximum 4TB
		LUSE	
7	Duplicated writing mode	Synchronized (Sync)	
8	Combination of the CUs	One-to-one correspondence	
9	Fence level	Data, Status, Never	Supports a function equivalent to the TC.
10	Multiple CU support	Yes	For CU#0 through CU#3F, TrueCopy pairs can be created.

Copyright © 2012, Hitachi, Ltd.

(1) Means for setting the paired LU:
The following three means are provided:
- RAID Manager/TrueCopy
- SVP
- Storage Navigator

Not only the pairing but also a series of pair state changes are possible by using these three means. However, the user can use two means only: the command instruction from the RAID Manager/TrueCopy and the instruction from the Web Console.

(2) LU size capable of the duplicated writing:
If two LUSE volumes are paired with TrueCopy, a LUSE P-VOL must be paired with S-VOL of the same size and the same structure. For example, if a LUSE P-VOL is connected with the volumes of 1GB, 2GB, and 3GB in this order, you must specify the LUSE volume which has exactly the same size and the same connection order as the S-VOL.

(3) Fence level:
The TrueCopy supports three types of fence level: Data, Status, and Never.

Copyright © 2012, Hitachi, Ltd.

◆ Restrictions:

(1) Command device:

① The TrueCopy provides users with a command to enable a state change and status display of the TrueCopy pair from the server.

② Assign a special LUN called a command device so that the DKC710I can receive this pair state change and pair status display commands.

③ Users cannot use the command device. The capacity allocated to command devices in a storage system is not available for users. The minimum capacity of command device is 47MB.

④ Use Web Console to specify the command device.

(2) Flashing updated data in the server:
When the TrueCopy is used as a DataPlex function, split the primary/secondary paired VOL. A Sync command or the like must be issued before splitting it and a file system buffer must be flashed when acquiring a backup from the secondary VOL. Thus, the latest backup can be acquired.

(3) P-VOL (primary VOL) access:
Pair suspend operation (pairsplit-P option) from RAID Manager/TrueCopy can be executed to TrueCopy pair volumes.

Copyright © 2012, Hitachi, Ltd.

3.23.7.3 Basic UR Specifications

Basic UR specifications are shown below.

Table 3.23.7.3-1 Basic Specifications of UR

No.	Item	Description	Remarks
1	Host interface on open system side	Fibre Channel	Conforming to base platform function.
2	Supporting OS platform	Conforming to base platform function.	
3	Connection between the CUs	Fibre Channel	
4	Means for setting the Journal volume	RAID Manager	
		Storage Navigator	
		Web Console	
5	Number of setting Journal volumes in a Journal Group	64 volumes	
6	LU type for setting the Journal volume	Normal VOL	
7	Means for setting the paired LU	RAID Manager/UR	
		Storage Navigator	
		Web Console	
8	Number of LUs capable of the duplicated writing	Maximum 16,382 pairs (*1)	
9	LU size capable of the duplicated writing (The paired VOL must be the same DEV type.)	Normal VOL	Maximum 3TB
		DP-VOL	Maximum 4TB
		LUSE	
10	Duplicated writing mode	Asynchronized (Async)	
11	Combination of the CUs	One-to-one correspondence	
12	Multiple DKC support	Yes	For CU#0 through CU#3F, UR pairs can be created.

*1: The number of maximum pairs is different depending on the volume size of each pair. Please refer to “Hitachi Universal Replicator User Guide” for the number of maximum pairs.

Copyright © 2012, Hitachi, Ltd.

3.23.7.3.1 Main and Remote Control Units (Primary storage systems and Secondary storage systems)

The main control unit (primary storage system) and remote control unit (secondary storage system) control UR operations:

- The primary storage system is the control unit in the primary storage system which controls the primary data volume of the UR pairs and master journal volume. The Storage Navigator Web Console PC must be LAN-attached to the primary storage system. The primary storage system communicates with the secondary storage system via the dedicated remote copy connections. The primary storage system controls the host I/O operations to the UR primary data volume and the journal obtain operation of the master journal volume as well as the UR initial copy and update copy operations between the primary data volumes and the secondary data volumes.

- The secondary storage system is the control unit in the secondary storage system which controls the secondary data volume of the UR pairs and restore journal volume. The secondary storage system controls copying of journals and restoring of journals to secondary data volumes. The secondary storage system assists in managing the UR pair status and configuration (e.g., rejects write I/Os to the UR secondary data volumes). The secondary storage system issues the read journal command to the primary storage system and executes copying of journals. The secondary Storage Navigator PC should be connected to the secondary storage systems at the secondary site on a separate LAN. The secondary storage systems should also be attached to a host system to allow sense information to be reported in case of a problem with a secondary data volume or secondary storage system and to provide disaster recovery capabilities.

3.23.7.3.2 Journal Group

Journal group consists of two or more data volumes and journal volumes. It is a feature that allows you to sort multiple data volumes and journal volumes into collective units to tailor UR to meet your unique business needs. The journal group in the primary storage system is referred to as the master journal group. The journal group in the secondary storage system is referred to as the restore journal group. The data volumes in the master journal group are also called the primary data volumes. The journal volumes in the master journal group are called the master journal volumes. The data volumes in the restore journal group are similarly called the secondary data volumes. The journal volumes in the restore journal group are called the restore journal volumes.
The data update sequence from the host is managed per the journal group. The data update sequence consistency between the master and restore journal groups to be paired is maintained and ensured. The master and restore journal groups are managed according to the journal group number. The journal numbers of master and restore journal groups that are paired can be different. One data volume and one journal volume can belong to only one journal group.

3.23.7.3.3 Data Volume Pair

UR performs remote copy operations for data volume pairs created by the user. Each UR pair consists of one primary data volume and one secondary data volume which can be located in different storage systems. The UR primary data volumes are the primary volumes (LDEVs) which contain the original data, and the UR secondary data volumes are the secondary volumes (LDEVs) which contain the backup or duplicate data. During normal UR operations, the primary data volume remains available to all hosts at all times for read and write I/O operations. During normal UR operations, the secondary storage system rejects all host-requested write I/Os for the secondary data volume. The secondary data volume write enable option allows write access to a secondary data volume while the pair is split and uses the secondary data volume and primary data volume track maps to resynchronize the pair (see section 3.23.7.4.4).

3.23.7.3.4 Journal Volume

When UR is used, updates to primary data volumes can be stored in other volumes, which are called journal volumes. The updates (which is sometimes referred to as update data) that will be stored in journal volumes are called journal data.
Because journal data will be stored in journal volumes, you can perform and manage highly reliable remote copy operations without suspension of remote copy operations. For example:

– Even if a communication path between the primary storage system and the secondary storage system fails temporarily, remote copy operations can continue after the communication path is recovered.

– If data transfer from hosts to the primary storage system is temporarily faster than data transfer between the primary storage system and the secondary storage system, remote copy operations between the primary storage system and the secondary storage system can continue. Because journal volumes can contain a lot more update data than the cache memory can contain, remote copy operations can continue if data transfer from hosts to the primary storage system is faster for a relatively long period of time than data transfer between the primary storage system and the secondary storage system.

(1) The Number of Journal Volumes
One journal group can contain up to 64 journal volumes. Each of the journal volumes can have different volume sizes and different RAID configurations. Journal data will be stored sequentially and separately into each journal volume in the same journal group.

Copyright © 2012, Hitachi, Ltd.

(2) Specifications of Journal Volumes

- Volumes and their capacity:
  You can use Virtual LUN (VLL) volumes for journal volumes.
  Journal volumes in the same journal group can be of different capacity. A master journal volume and the corresponding restore journal volume can be of different capacity.
  A journal volume consists of two areas: one area is used for storing journal data, and the other area is used for storing metadata for remote copy.

- RAID configuration:
  Journal volumes support all RAID configurations that are supported by DW700. Journal volumes also support all physical volumes that are supported by DW700.

- Support for option programs:
  Cache Residency Manager volumes can be used for journal volumes.

Copyright © 2012, Hitachi, Ltd.

(3) Restrictions on Journal Volumes

- Registering journal volumes:

  NOTE: You must register journal volumes in a journal group before you create a data volume pair for the first time in the journal group.

  You can add journal volumes under any of the following conditions:
  – When the journal group does not contain data volumes (i.e., before you create a data volume pair for the first time in the journal group, or after all data volume pairs are deleted)
  – When all data volume pairs in the journal group are suspended
  – When processing for changing the status of a data volume pair (for example, deletion or suspension of a data volume pair) is not in progress

  NOTE: If a path is defined from a host to a volume, you cannot register the volume as a journal volume.

  You can use Storage Navigator computers to register journal volumes.
  If you add a journal volume when a remote copy operation is in progress (i.e., when at least one data volume pair exists for data copying), the metadata area of the journal volume (see (4)) will be unused and only the journal data area will be used. To make the metadata area usable, you need to split (suspend) all the data volume pairs in the journal group and then restore (resynchronize) the pairs.
  Adding journal volumes during a remote copy operation will not decrease the metadata usage rate if the metadata usage rate is high.
  Adding journal volumes during a remote copy operation may not change the journal data usage rate until the journal volumes are used. To check the journal data usage rate, use the Usage Monitor panel.

- Deleting journal volumes:
  You can delete journal volumes under any of the following conditions:
  – When the journal group does not contain data volumes (i.e., before you create a data volume pair for the first time in the journal group, or after all data volume pairs are deleted)
  – When all data volume pairs in the journal group are suspended
  You can use Storage Navigator computers to delete journal volumes.

- Access from hosts to journal volumes:
  If a path is defined from a host to a volume, you cannot register the volume as a journal volume.
  You cannot define paths from hosts to journal volumes. This means that hosts cannot read from and write to journal volumes.

Copyright © 2012, Hitachi, Ltd.

(4) Journal Volume Areas

The journal volume consists of the metadata area and the journal data area. The ratio of metadata area to journal data area is common in the journal volumes within the journal group.

In the metadata area, the metadata that manages the journal data is stored. For further information on the metadata area, see Table 3.23.7.5.1-1. The journal data that the metadata manages is stored in the journal data area.

Copyright © 2012, Hitachi, Ltd.

3.23.7.3.5 Remote Copy Connections

The remote copy connections are the physical paths used by the primary storage systems to communicate with the secondary storage systems. Remote copy connections enable communication between the primary and secondary storage systems. The primary storage systems and secondary storage systems are connected via fibre-channel interface cables. You must establish paths from the primary to the secondary storage system, and also from the secondary to the primary storage system. Up to eight paths can be established in both of these directions.

When fibre-channel interface (optical multimode shortwave) connections are used, two switches are required for distances greater than 0.5 km (1,640 feet), and distances up to 1.5 km (4,920 feet, 0.93 miles) are supported. If the distance between the primary and secondary sites is greater than 1.5 km, the optical single mode longwave interface connections are required. When fibre-channel interface (single-mode longwave) connections are used, two switches are required for distances greater than 10 km (6.2 miles), and distances up to 30 km (18.6 miles) are supported.

 Copyright © 2012, Hitachi, Ltd.

3.23.7.3.6 Initiator Ports and RCU Target Ports

The initiator port and the RCU target port are required at both the primary storage system and secondary storage system. The initiator port at the primary storage system is connected to the RCU target port at the secondary storage system via the fibre channel interface. The initiator port at the secondary storage system is connected to the RCU target port at the primary storage system. The initiator port at the secondary storage system issues a “read journal” command to the primary storage system, and then the RCU target port at the primary storage system sends journal data to the secondary storage system in response to the “read journal” command.

Any fibre-channel interface port of the DW700 can be configured as an initiator port. The initiator ports cannot communicate with the host processor channels. The host channel paths must be connected to the fibre-channel interface port other than the initiator port.

3.23.7.3.7 UR Web Console Software

DW700 Storage Navigator Java applet program product includes UR for the DW700 storage system. The DW700 Storage Navigator software communicates with the SVP of each DW700 storage system via defined TCP/IP connections.

The Storage Navigator PC at the primary site must be attached to the primary storage system. You should also attach a Storage Navigator PC at the secondary site to all secondary storage systems. Having a Storage Navigator PC at the secondary site enables you to change the UR parameter of the secondary storage system and access the UR secondary data volume (e.g. for the maintenance of media). If you need to perform UR operations in the reverse direction from the secondary site to the primary site (e.g., disaster recovery), the DW700 UR software simplifies and expedites this process.

Copyright © 2012, Hitachi, Ltd.

3.23.7.4 Remote Copy Operations

Figure 3.23.7.4-1 illustrates the two types of UR remote copy operations: initial copy and update copy.

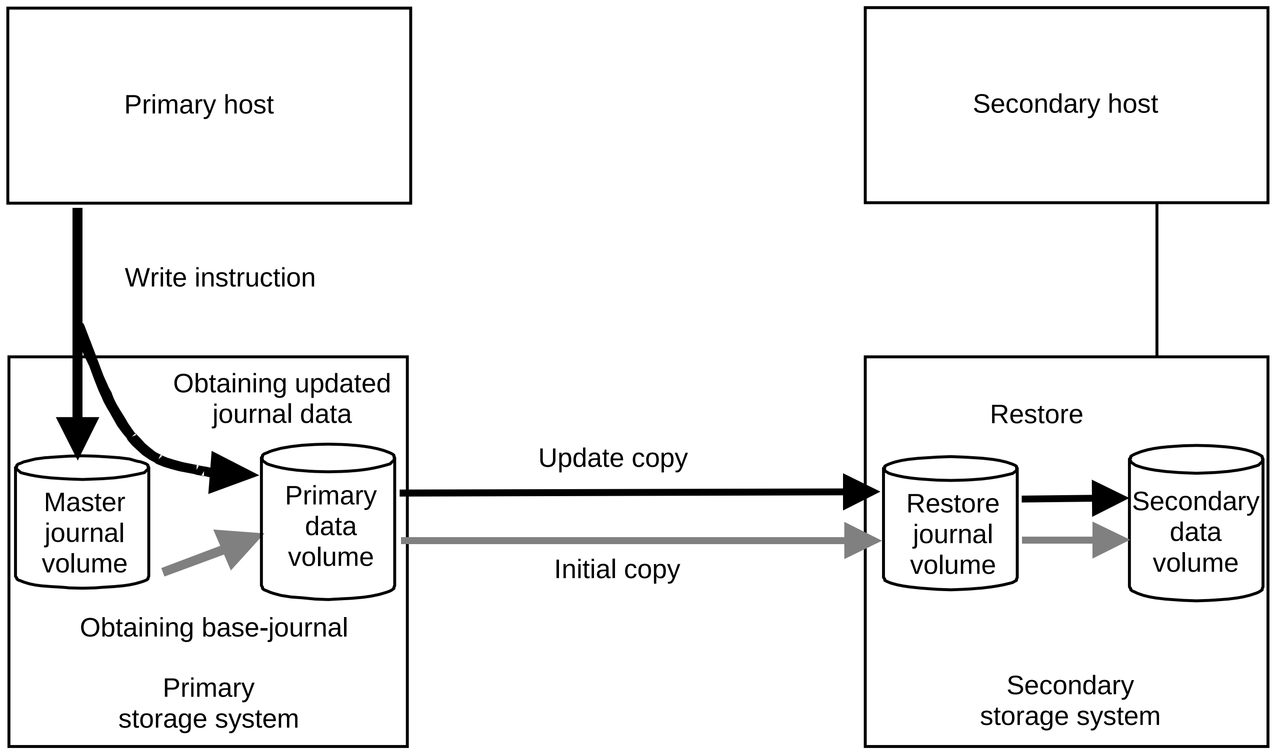


Figure 3.23.7.4-1 Remote copy operations

This section describes the following topics that are related to remote copy operations with UR:

- Initial copy operation (see section 3.23.7.4.1)
- Update copy operation (see section 3.23.7.4.2)
- Read and write I/O operations during UR volumes (see section 3.23.7.4.3)
- Secondary data volume write option (see section 3.23.7.4.4)
- Difference management (see section 3.23.7.4.5)

 Copyright © 2012, Hitachi, Ltd.

3.23.7.4.1 Initial Copy Operation

Initial copy operations synchronize data in the primary data volume and data in the secondary data volume. Initial copy operations are performed independently from host I/Os. Initial copy operations are performed when you create a data volume pair or when you resynchronize a suspended pair. The initial copy operation copies the base-journal data that is obtained from the primary data volume at the primary storage system to the secondary storage system, and then restores the base-journal to the secondary data volume.

If the journal-obtain operation starts at the primary data volume, the primary storage system obtains all data of the primary data volume as the base-journal data, in sequence. The base-journal contains a replica of the entire data volume or a replica of updates to the data volume. The base-journal will be copied from the primary storage system to the secondary storage system after the secondary storage system issues a read-journal command. After a base-journal is copied to the secondary storage system, the base-journal will be stored in a restore journal volume in a restore journal group where the secondary data volume belongs. After that, the data in the restore journal volume will be restored to the secondary data volume, so that the data in the secondary data volume synchronizes with the data in the primary data volume.
The base-journal data is stored in the entire data volume or the area for the difference. The area for the difference is used when the difference resynchronization operation is performed. The journal data for the entire data volume is created when the data volume pair is created. The difference journal data is obtained when the pair status of the data volume changes from the Suspending status to the Pair resync status. Merging the difference bitmaps that are recorded on both primary and secondary data volumes enables you to obtain the journal data for only difference. When a data volume pair is suspended, the status of data that is updated from the host to the primary and secondary data volumes is recorded to the difference bitmap.
The base-journal data of primary storage system is stored to the secondary storage system journal volume according to the read command from the secondary storage system. After that, the base-journal data is restored from the journal volume to the secondary data volume. The initial copy operation will finish when all base-journals are restored.

NOTE: If you manipulate volumes (not journal groups) to create or resynchronize two or more data volume pairs within the same journal group, the base journal of one of the pairs will be stored in the restore journal volume, and then the base journal of another pair will be stored in the restore journal volume. Therefore, the operation for restoring the latter base journal will be delayed.

NOTE: You can specify None as the copy mode for initial copy operations. If the None mode is selected, initial copy operations will not be performed. The None mode must be used at your responsibility only when you are sure that data in the primary data volume is completely the same as data in the secondary data volumes.

Copyright © 2012, Hitachi, Ltd.

3.23.7.4.2 Update Copy Operation

When a host performs a write I/O operation to a primary data volume of a data volume pair, an update copy operation will be performed. During an update copy operation, the update data that is written to the primary data volume is obtained as an update journal. The update journal will be copied to the secondary storage system, and then restored to the secondary data volume.

The primary storage system obtains update data that the host writes to the primary data volume as update journals. Update journals will be stored in journal volumes in the journal group that the primary data volume belongs to. When the secondary storage system issues “read journal” commands, update journals will be copied from the primary storage system to the secondary storage system asynchronously with completion of write I/Os by the host. Update journals that are copied to the secondary storage system will be stored in journal volumes in the journal group that the secondary data volume belongs to. The secondary storage system will restore the update journals to the secondary data volumes in the order write I/Os are made, so that the secondary data volumes will be updated just like the primary data volumes are updated.

Copyright © 2012, Hitachi, Ltd.

3.23.7.4.3 Read and Write I/O Operations During UR Volumes

When a primary storage system receives a read I/O for a UR primary data volume, the primary storage system performs the read from the primary data volume. If the read fails, the redundancy provided by RAID1, RAID5 or RAID6 technology recovers the failure. The primary storage system does not read the UR secondary data volume for recovery.
When a primary storage system receives a write I/O for the primary data volume with PAIR status, the primary storage system performs the update copy operation, as well as writing to the primary data volume.
The primary storage system completes the primary data volume write operations independently of the update copy operations at the secondary data volume. The secondary storage system updates the data in the secondary data volume according to the write sequence number of journal data. This will maintain the data consistency between the primary and secondary data volumes. If the primary data volume write operation fails, the primary storage system reports a unit check and does not create the journal data for this operation. If the update copy operation fails, the secondary storage system suspends either the affected pair or all UR pairs in the journal group, depending on the type of failure. When the suspended UR pair or journal group is resumed (Resume Pair), the primary storage system and secondary storage system negotiate the resynchronization of the pair(s).
During normal UR operations, the secondary storage system does not allow UR secondary data volumes to be online (mounted), and therefore hosts cannot read from and write to secondary data volumes. The UR secondary data volume write enable option allows write access to a secondary data volume while the pair is split (see section 3.23.7.4.4). The secondary data volume write option can only be enabled when you split the pair from the primary storage system.

Copyright © 2012, Hitachi, Ltd.

3.23.7.4.4 Secondary Data Volume Write Option

For additional flexibility, UR provides a secondary data volume write option (S-Vol. Write) which enables write I/O to the secondary data volume of a split UR pair. The secondary data volume write option can be selected by the user during the Suspend Pair operation and applies only to the selected pair(s). The secondary data volume write option can be accessed only when you are connected to the primary storage system. When you resync a split UR pair which has the secondary data volume write option enabled, the secondary storage system sends the secondary data volume track bitmap to the primary storage system, and the primary storage system merges the primary data volume and secondary data volume bitmaps to determine which tracks are out-of sync. This ensures proper resynchronization of the pair.

3.23.7.4.5 Difference Management

The differential data (updated by write I/Os during split or suspension) between the primary data volume and the secondary data volume is stored in each track bitmap. When a split/suspended pair is resumed (Resume Pair), the primary storage system merges the primary data volume and secondary data volume bitmaps, and the differential data is copied to the secondary data volume.

Copyright © 2012, Hitachi, Ltd.

3.23.7.5 Journal Processing

The UR journal data contains the primary data volume updates and the metadata information (associated control information), which enables the secondary storage system to maintain update consistency of the UR secondary data volumes. UR journal processing includes:

- Creating and storing journals at the primary storage system (see section 3.23.7.5.1)
- Copying journals to the secondary storage system (see section 3.23.7.5.2)
- Storing journals at the secondary storage system (see section 3.23.7.5.3)
- Selecting and restoring journals at the secondary storage system (see section 3.23.7.5.4)
- Types of journals (see section 3.23.7.5.5)

Copyright © 2012, Hitachi, Ltd.

3.23.7.5.1 Creating and Storing Journals at the Primary storage system

When a primary storage system performs an update (host-requested write I/O) on a UR primary data volume, the primary storage system creates a journal data to be transferred to secondary storage system. The journal data will be stored into the cache at first, and then into the journal volume.
Metadata information will be attached to journal data (see Table 3.23.7.5.1-1). When base-journal is obtained, only metadata information is created and stored in UR cache or the journal volume.

Table 3.23.7.5.1-1 Metadata Information

Type	Description
Journal type	Type of journal (e.g., base-journal or update journal)
LDEV No. (data)	The number of primary data volume that stores the original data
Original data storing position	The primary data volume slot number, and the start and end of sub-block number (data length)
LDEV No. (journal)	The volume number of master journal volume that stores the journal data
Journal data storing position	The slot number of master journal volume, and the start sub-block number
Journal sequence number	The sequence number that is assigned when the journal is obtained

The journal sequence number indicates the primary data volume write sequence that the primary storage system has created for each journal group. The journal data is transferred to the secondary storage system asynchronously with the host I/O. The secondary storage system updates the secondary data volume in the same order as the primary data volume according to the sequence number information in the journal.

3.23.7.5.2 Copying Journals to the Secondary storage system

When a primary storage system receives a read journal command from a secondary storage system, the primary storage system sends the journal data to the secondary storage system. The secondary storage system's initiator ports act as host processor channels and issue special I/O operations, called remote I/Os (RIOs), to the primary storage system. The RIO transfers the journal data in FBA format using a single channel command. The primary storage system can send several journal data using a single RIO, even if their sequence numbers are not contiguous. Therefore, the journal data are usually sent to the secondary storage system in a different order than the journal data were created at the primary storage system. The secondary storage system ensures that the journal data are applied to the secondary data volume in the correct sequence. This method of remote I/O provides the most efficient use of primary storage system-to-secondary storage system link resources.

3.23.7.5.3 Storing Journal at the Secondary storage system

A secondary storage system receives the journal data that is transferred from a primary storage system according to the read journal command. The journal data will be stored into the cache at first, and then into the journal volume.

NOTE: The primary storage system does not remove the target journal data from its master journal volume until it receives the sequence numbers of restored journal which is give to the read journal command from the secondary storage system. This is true even if the primary storage system and secondary storage system are connected via a channel extender product.

Copyright © 2012, Hitachi, Ltd.

3.23.7.5.4 Selecting and Restoring Journal at the Secondary storage system

The secondary storage system selects journal data to be promoted to formal data (or “restored”) as follows:

1. The secondary storage system gives the number as the management information to distinguish the journal data arrival to the sequence number that is assigned to the journal data from the primary storage system. If the number is 1, the journal data arrived at the secondary storage system. If the number is 0, the journal data has not arrived yet. The secondary storage system determines whether the journal data should be settled or not according to this number. If the journal data has not arrived yet, the secondary storage system waits for the journal data.
2. When the top of queue in the journal group indicates the journal data arrival, the secondary storage system selects the journal data which has the lowest sequence number, and then settles this journal data.
3. The secondary storage system repeats steps 1. and 2. to select and settle the journal data.

Figure 3.23.7.5.4-1 illustrates the journal data selection and settling at the secondary storage system. This diagram shows that journal data S1 arrives at the secondary storage system because the management information indicates 1. The secondary storage system selects journal data S1 to be settled, because S1 is the lowest sequence number. When S1 is removed from the queue of sequence numbers, journal data S2 becomes the top entry, but it has not arrived yet. The management information of journal data S2 is 0. The secondary storage system waits journal data S2. When journal data S2 arrives, the secondary storage system selects S2 as the next journal data to be settled. The journal data selected by the secondary storage system is marked as “host-dirty” and treated as formal data.

Copyright © 2012, Hitachi, Ltd.

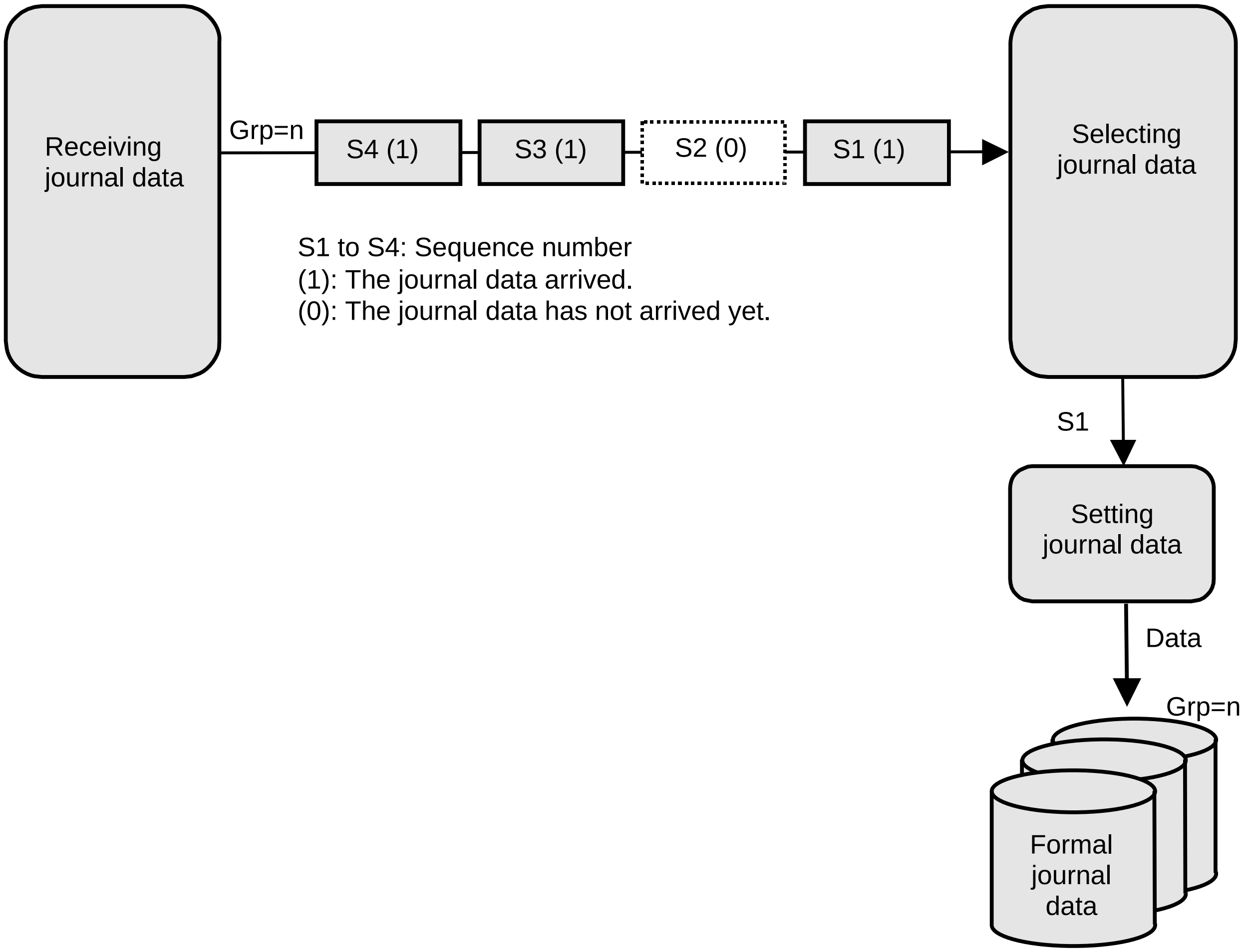


Figure 3.23.7.5.4-1 Selecting and Settling Journal at the Secondary storage system

The secondary storage system settles and restores the journal data to the secondary data volume as follows:

- Journal data stored in the cache
  The journal data is copied to the corresponding cached track and promoted to formal data.
- Journal data stored in the restore journal volume
  The journal data is read from the restore journal volume to cache. The journal data that is read to cache is copied to the existing cache track and promoted to formal data. After that, the space for the restore journal volume is released.

Copyright © 2012, Hitachi, Ltd.

3.23.7.5.5 Types of Journal

In addition to the journal data for updating, the primary subsystem sends control information to the secondary subsystem. This control information indicates when volume pair status changes and when a primary subsystem power-off sequence is initiated.

Copyright © 2012, Hitachi, Ltd.

3.23.7.6 UR operation

3.23.7.6.1 Pair operation

The following figure illustrates an UR pair configuration. In the configuration, UR pairs belong to the journal group. Each journal group and UR pair have an attribute and status. Each attribute and status is described in the following subsections.

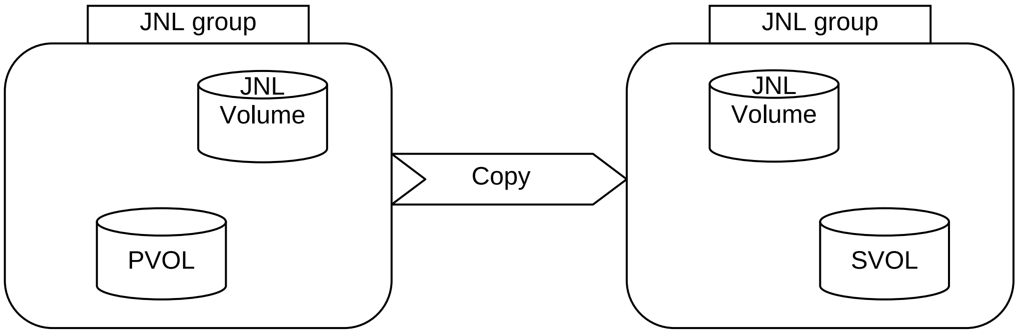


Figure 3.23.7.6.1-1 Relationship between pairs and journal groups

(1) Journal group
Prior to pair creation, the journal volume needs to be registered, and the journal group needs to be defined. You can delete journal groups that are not in use. To define journal groups, you need to register journal volumes. On the other hand, to delete journal groups, you need to delete all journal volumes.

Table 3.23.7.6.1-1 Journal group attributes

#	Journal group attribute	Meaning
1	Initial	No UR pair is set in the journal group to which journal volumes are registered.
2	Master	The journal group to which the logical volume storing the original data (PVOL) belongs. The attribute of the journal group in Master is called “M journal”.
3	Restore	The journal group to which the logical volume storing the duplicated data (SVOL) belongs. The attribute of the journal group in Restore is called “R journal”.

Table 3.23.7.6.1-2 Journal group status

#	Journal group status	Meaning
1	–	UR pairs are not set in the target group
2	Active	Base/update copy is in progress in the target group
3	Halting	Base/update copy is halted in the target group
4	Stopping	The target group is being suspended, or being deleted
5	Stop	The target group is suspended. (All UR pairs in the target group are suspended)

Copyright © 2012, Hitachi, Ltd.

Table 3.23.7.6.1-3 Journal group operations

#	Operation	Specified attribute	Description
1	Register journal volume	MJNL/RJNL	✓ Register a logical volume, to which no path is defined, as a JNL volume. ✓ Can register the additional volume to the existing journal group only when the UR remote copy is stopped (Journal group status is STOP). ✓ When the journal group state is ACTIVE, the journal volume can be registered. (However, only the journal area increases, and after the journal group suspend and resync, the meta data area increases.) ✓ Can register UR pair to the journal group to which the journal volume is registered.
2	Delete journal volume	MJNL/RJNL	✓ Delete the logical volume registered as a journal volume from the journal volume. ✓ Can delete it only when UR remote copy is stopped (Journal group status is STOP or the group attribute is Initial) if there are two or more journal volumes in the target journal group. ✓ Can delete it only when the target group does not have any UR pair (Group attribute is Initial) if the number of journal volumes in the target journal group is one. ✓ No UR pair can be registered to such a journal group from which all journal volumes are deleted.

Copyright © 2012, Hitachi, Ltd.

(2) UR pairs

UR performs a remote copy operation for the logical volume pair set by the user. Based on pair operations, pair attributes and pair statuses are added to the logical volumes. You can perform the following remote copy operations for UR pairs. The figure below illustrates how the UR pair status changes due to each operation.

Table 3.23.7.6.1-4 Pair attributes

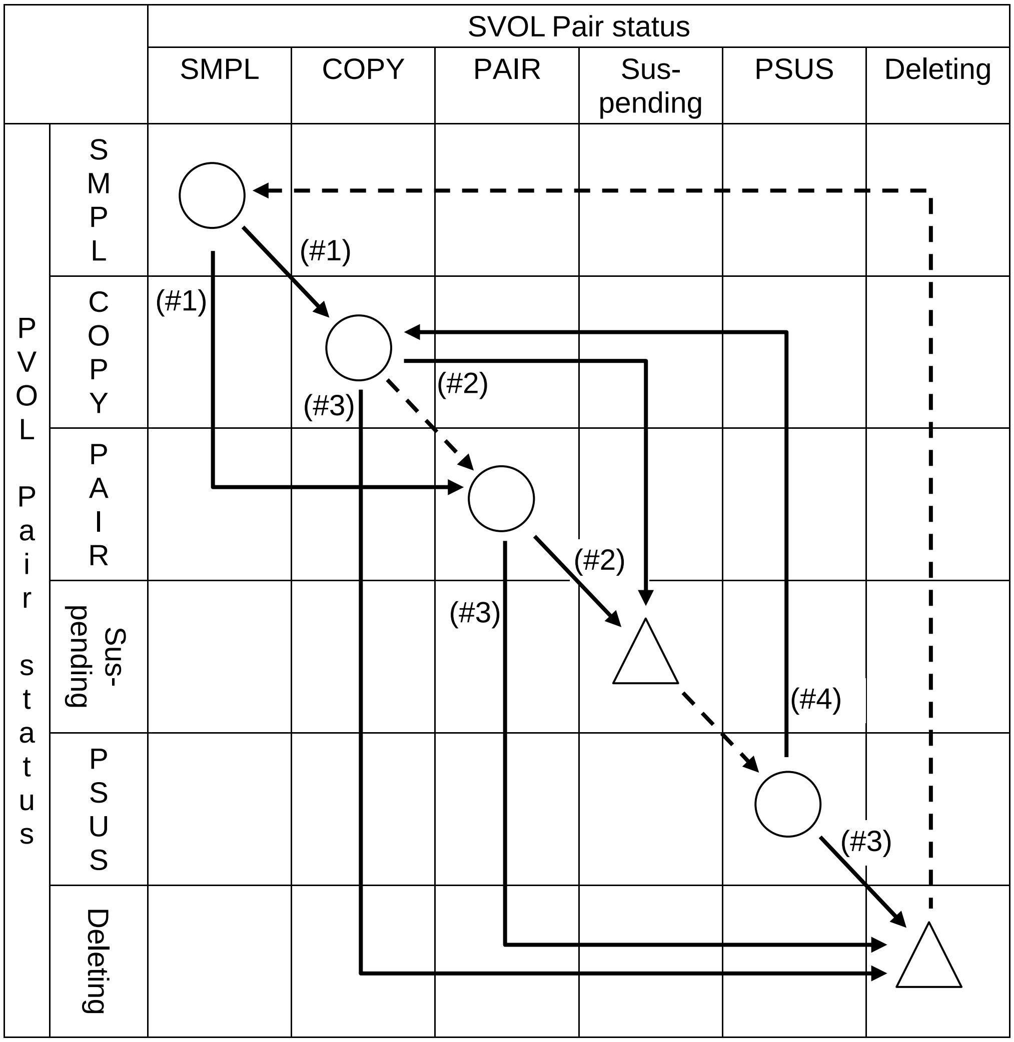
#	Pair attribute	Description
1	SMPL	Target volume is not assigned to an UR pair.
2	PVOL	Primary volume. Data volume to which the original data is stored.
3	SVOL	Secondary volume. Data volume to which backup or duplicated data is stored.

Table 3.23.7.6.1-5 Pair status

#	Pair status	Description
1	SMPL	Target volume is not assigned to an UR pair.
2	COPY	Base copy is in progress and data of the PVOL and SVOL of the UR pair do not match completely. When their data match, the status changes to PAIR.
3	PAIR	Base copy is completed, and data of the PVOL and SVOL match completely.
4	PSUS/SSUS	Copy operation is suspended in the UR pair.
5	PSUE	An error is detected in the DKC, and the copy in the UR pair is stopped (Suspended).
6	PFUL	The journal usage in the journal volume exceeded the threshold. The copy operation is continued.
7	PFUS	The capacity of the stored journal exceeded the journal volume capacity, and the copy operation is suspended in DKC.
8	SSWS	Data can be written to the SVOL in which Takeover is in progress.
9	Suspending	The status of the UR pair is being changed to Suspend.
10	Deleting	The UR pair is deleted, and the status is being changed to SMPL.

Table 3.23.7.6.1-6 Pair operations

#	Operation	Specified attribute	Description
1	Pair create	MJNL	Register the logical volume to which a path is defined as an UR pair. There are two types of copy instruction, “All copy” and “NO copy”. “All copy” performs a base copy, and “NO copy” does not perform a base copy.
2	Pair suspend	MJNL/RJNL	Change the status of the UR pair, which is performing the base/update copy, to the suspend status.
3	Pair delete	MJNL/RJNL	Delete the already registered UR pair.
4	Pair resync	MJNL/RJNL	Change the pair status of the UR, in which the copy operation is suspended, to the pair resume status. (RJNL can be specified only when swapping is specified)
5	Takeover	RJNL	Swap MJNL and RJNL (reverse the source and the target) and resync the pair.



(#x) : Correspond to the pair operation # in Table 3.23.9.6.1-6.
→ : Status transition due to pair operation
- - - -→ : Status transition in DKC
△ : Status transition is in progress. DKC internal processing is in progress.
○ : Normal status. Status is changed to this when operation completes.

Figure 3.23.7.6.1-2 Pair attributes

Copyright © 2012, Hitachi, Ltd.

(3) Pair create

Logical volume to which a path is defined is registered to the target journal group as an UR pair (data volume).

- Pair create option

  When creating a pair, you can set a pair create option. The following table shows the options and applications that can be specified.

Table 3.23.7.6.1-7 Pair create options

#	Option	Feature overview	SN	RM
1	Copy type	You can choose to or not to perform a base copy. The following copy types are available. ✓ All copy: Perform base copy when creating a pair ✓ No copy: Not perform base copy when creating a pair	○	○
2	Priority	Specify which pair you want to perform a base copy first when creating multiple pairs at a time. ✓ Setting: 1-255	○	–
3	Error level	Set error levels. You can choose either of the following levels. ✓ Group: Even if an error which affects only the specified volume occurs, all pairs in the journal group will suspend due to the error. ✓ Volume: When an error which affects only the specified volume occurs, only that pair suspends due to the error. However, when an error which affects the entire journal group occurs, all pairs in the journal group are suspended.	○	–

Legend) SN: Instruction from Storage Navigator (OPEN), RM: Instruction from RAID Manager

Copyright © 2012, Hitachi, Ltd.

(4) Pair suspend

The copy operation of an UR pair is stopped, and the pair status is changed to PSUS. When all UR pairs in a journal group are suspended, the status of the journal group is changed to STOP. The specifications of the suspend operation are as follows.

- ✓ This operation is performed when the target volume is in the COPY/PAIR status and all volumes in the journal group are in the status other than Suspending/Deleting status.
- ✓ The pair suspend operation can be performed from PVOL/SVOL. The processing of the suspend operation is the same for both PVOL/SVOL instructions.
- ✓ When you perform the pair suspend operation, you can specify the Pend Update and the suspend range. The table below shows the relationships of Pend Update and the suspend range.

Table 3.23.7.6.1-8 Pair suspend

#	Pend Update	Feature overview	Volume	Group
1	Flush	✓ When the suspend operation is received, the pending data in MCU/RCU is reflected. ✓ When the operation is performed while the host I/O is being processed, the contents of the PVOL and SVOP are not the same. (In the operation after the host I/O is stopped, the contents of the PVOL and SVOL are the same)	○	○
2	Purge	✓ When the suspend operation is received, the difference of the pending data in MCU/RCU is recorded to the differential bitmaps. (Since the data is not reflected, the contents of the PVOL and SVOL are not the same) ✓ The status can be changed to Suspend in a short time.	–	○

Copyright © 2012, Hitachi, Ltd.

- Pair suspend option

  When you perform a pair suspend operation, you can specify a pair suspend option. The following table shows the options and applications you can specify.

Table 3.23.7.6.1-9 Pair suspend option

#	Option	Feature overview	SN	RM
1	Pend Update	You can choose a Pend Update type when performing a pair suspend. ✓ Flush: The Pend data is processed in the Flush mode. ✓ Purge: The Pend data is processed in the Purge mode.	○	○
2	Range	You can specify the suspend range. (However, when the Purge mode is selected in Pend Update, only “Group” can be selected) ✓ Volume: Only the specified UR pairs are suspended. ✓ Group: All pairs in the specified journal group are suspended.	○	○
3	SVOL write	You can choose to perform Write/Read to the SVOL when performing a pair suspend. ✓ Disable: Cannot Read/Write data to the suspended SVOL. ✓ Enable: Can Read/Write data to the suspended SVOL.	○	○

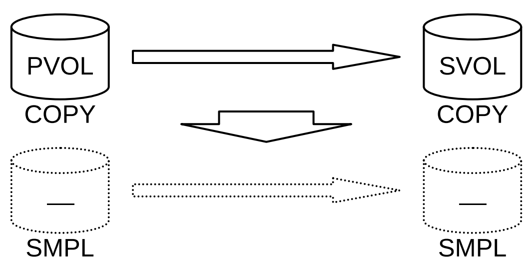
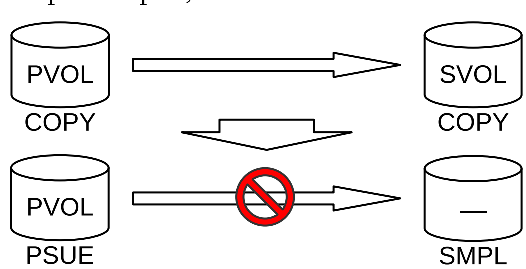
 Copyright © 2012, Hitachi, Ltd.

(5) Pair delete

The UR pair is deleted, and the pair volumes are restored to non-UR logical volumes. When all UR pairs in the target journal group are deleted, the attribute of the journal group is changed to Initial. The following table shows the specifications of the pair delete operation.

- ✓ This operation is performed when the target volume is in the normal status (COPY/PAIR/PSUS(PSUE)) and all volumes in the journal group are in the status other than Suspending/Deleting.
- ✓ If you perform the pair delete operation for an UR pair after stopping the host I/O, the data in the PVOL and that in the SVOL are the same. (When this operation is performed while the host I/O is in progress, the data are not the same.)
- ✓ The pair delete operation can be performed from PVOL/SVOL. Note that the feature depends on the specified volume attribute. The following table shows the specified volume attributes and the features.

Table 3.23.7.6.1-10 Pair delete

#	Specified volume	Feature overview
1	PVOL specified	✓ The statuses of both PVOL and SVOL are changed to SMPL, and the UR pair will be deleted. 
2	SVOL specified	✓ Only the status of SVOL changes to SMPL, and the PVOL status is not changed. If the PVOL is in the COPY/PAIR status, the pair will be suspended due to a failure (PSUE). When the operation is performed to the suspended pair, the PVOL status is not changed.  ✓ The UR pair of only the PVOL after the pair delete operation cannot be resynchronized. Therefore, you need to delete the pair using the PVOL-specified pair delete operation.

Copyright © 2012, Hitachi, Ltd.

- Pair delete option

  When you perform a pair delete operation, you can specify a pair delete option. The following table shows the options and applications you can specify.

Table 3.23.7.6.1-11 Pair delete option

#	Option	Feature overview	SN	RM
1	Range	You can specify the following pair delete range. ✓ Volume: Only the specified UR pairs are deleted. ✓ Group: All pairs in the specified journal group are deleted.	○	○
2	Specified volume attribute	You can specify the attribute of the volume to which the pair delete is performed. PVOL specified: Normal pair delete operation is performed. SVOL specified: Pair is deleted only in SVOL.	○	○
3	ForceDelete	You can specify ForceDelete which deletes pairs regardless of the pair status. The overview of the ForceDelete feature is as follows. ✓ ForceDelete deletes only the pairs in the specified volume without recognizing the status of the paired volume. Therefore, the operations in the target (paired) volume is not guaranteed. ✓ Only “Group” range is available. (Note that only “Volume” range is available for SMPL volumes) ✓ After the ForceDelete operation, the data in PVOL and that in SVOL are not the same.	○	–

 Copyright © 2012, Hitachi, Ltd.

(6) Pair resync (Pair resume)
The copy operation of the suspended UR pair is resumed. When a copy operation of one pair is resumed in the journal group in which the copy operation is stopped (STOP), the journal group status will be changed to Active. The specifications of pair resync operation are as follows.

✓ This operation is performed when the target UR pair is in the PSUS/PSUE status and the journal group to which the target UR belongs does not contain UR pair in the Suspending/Deleting status. However, this operation cannot be performed to the PVOL which is paired with the SVOL in the SSWS status.

✓ The operations can be specified in PVOL. However, the Swap instruction can be specified in SVOL optionally.

● Pair resync option
When you perform a pair resync operation, you can specify a pair resync option. In addition, when the range of the operation is “volume”, you can change the pair create option for the target volume. The following table shows the options and applications you can specify.

Table 3.23.7.6.1-12 Pair resync option

#	Option	Feature overview	SN	RM
1	Range	You can specify the following pair resync range. ✓ Volume: Only the specified UR pairs are resynchronized. ✓ Group: All pairs in the specified journal group are resynchronized.	○	○
2	Priority	To resynchronize multiple pairs at a time, set the priority of the base copy operation. ✓ Range: 1-255	–	–
3	Error level	Set the error level in the case of a failure. (Enabled only when the range “volume”) ✓ Group: Even if a failure affects only the specified volume, all pairs in the journal group will be suspended due to the failure. ✓ Volume: When a failure affects only the specified volume, only the pair will be suspended due to the failure. However, when a failure affects the entire journal group, all pairs in the journal group will be suspended due to the failure.	○	–
4	Swap	This option is specified in the SVOL, the volumes are swapped (SVOL→PVOL/PVOL→SVOL). The differential data is resynchronized from the swapped PVOL to SVOL. See “3.23.7.6.1 (7) Takeover” for the swapping feature.	–	○

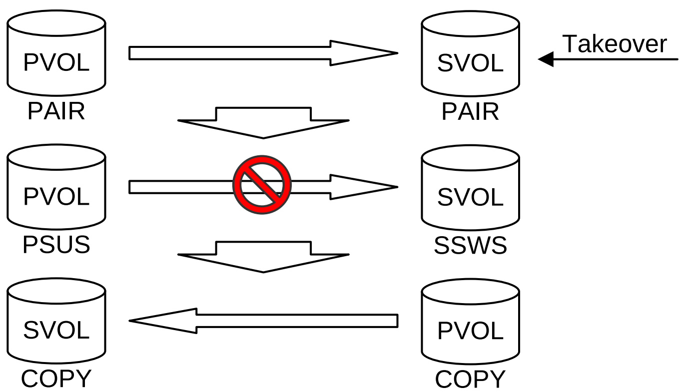
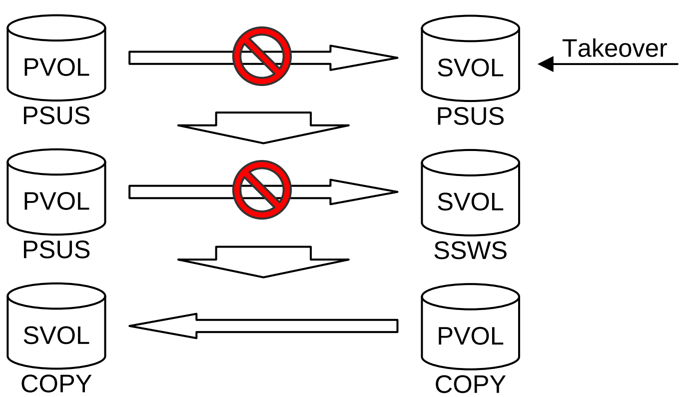
Copyright © 2012, Hitachi, Ltd.

(7) Takeover (Reverse Resync)

Takeover swaps PVOL and SVOL (switch the source and the target) and resynchronizes the pair. The specifications of the Takeover operation are as follows.

- ✓ Only RAID Manager can perform this operation.
- ✓ This operation is performed to SVOL. It can be performed even if the PVOL (i.e. DKC in the MCU) is lost due to disaster etc.
- ✓ This operation is performed when the pair status of the SVOL is PAIR or PSUS/SSWS and when the target RCU does not contain UR pairs in the Suspending/Deleting or COPY status. The following table shows when the Takeover command is issued.

Table 3.23.7.6.1-13 Takeover

#	Pair status	Feature overview
1	PAIR	(1) When a Takeover command is issued to the RCU, the Flush suspend is performed for the specified group. When it is issued, the SVOL status changes to SSWS, and it becomes a volume that can be read/written. (2) After the pair is suspended, a pair resync swap is performed to swap the PVOL and the SVOL, and an initial copy operation is performed. 
2	PSUS/PSUE	(1) When a Takeover command is issued to RCU, the Flush suspend is not performed for the specified group. Only the SVOL status is changed to SSWS. (2) After the suspend operation completes, a pair resync swap is performed to swap the PVOL and the SVOL, and an initial copy is performed. 

Copyright © 2012, Hitachi, Ltd.

(8) JNL group status and display of pair status by RAID Manager
We summarize the relationship between display of JNL group status and display of pair status regarding RAID Manager (RM) compared with JAVA display as the following.

① Operation suspend

#	JAVA Pair status	RM Pair status	RM (M-JNL) JNLGroup status	RM (R-JNL) JNLGroup status	Remarks
1	SMPL	SMPL	–	–	–
2	COPY/PAIR	COPY/PAIR	PJNN	SJNN	–
3	Suspending				–
4	PSUS	PSUS/SSUS	PJSN	SJSN	–

② JNL utilization threshold over (JNL utilization is over <u>80%</u> (but...not failure))

#	JAVA Pair status	RM Pair status	RM (M-JNL) JNLGroup status	RM (R-JNL) JNLGroup status	Remarks
1	SMPL	SMPL	–	–	–
2	COPY/PAIR	PFUL	PJNF	SJNF	–
3	Suspending	–	–	–	Pair status will not change to “Suspend”.
4	PSUS	PSUS/SSUS	–	–	Same as above

③ Failure suspend (except for JNL puncture)

#	JAVA Pair status	RM Pair status	RM (M-JNL) JNLGroup status	RM (R-JNL) JNLGroup status	Remarks
1	SMPL	SMPL	–	–	–
2	COPY/PAIR	COPY/PAIR	PJNN	SJNN	–
3	Suspending	PSUE	PJSE	SJSE	–
4	PSUE	PSUE			–

④ JNL puncture suspend

#	JAVA Pair status	RM Pair status	RM (M-JNL) JNLGroup status	RM (R-JNL) JNLGroup status	Remarks
1	SMPL	SMPL	–	–	–
2	COPY/PAIR	COPY/PAIR	PJNN	SJNN	–
3	Suspending	PFUS	PJSF	SJSF	–
4	PSUS	PFUS			–

3.23.7.6.2 USAGE/HISTORY

(1) USAGE
This feature enables you to view the UR information (Frequency of I/O for UR pair, transfer rate of journals between MCU/RCU, usage of journals in MCU/RCU etc.) using SVP and Web Console. The specifications are as follows.

Table 3.23.7.6.2-1 USAGE specification

#	Item	Specification etc.
1	Sampling interval	1-15 min. (1 min.)
2	Samplings	1440
3	Unit of sampling/display	• For each LU • For each JNL group • For each system
4	Sampling item	See Table 3.23.7.6.2-2
5	Others	You can save the monitor data in the text format using the Export Tool of Performance Monitor.

Copyright © 2012, Hitachi, Ltd.

Table 3.23.7.6.2-2 Sampled items

#		Category	Statistics	Description
1	MCU	Host I/O	Write Record Count	The number of Write I/Os per second
2			Write Transfer Rate	The amount of data that are written per second (KB/sec.)
3		Initial copy	Initial copy HIT rate	Initial copy hit rate (%)
4			Average Transfer Rate	The average transfer rate for initial copy operations (KB/sec.)
5	MCU	Async copy	M-JNL Asynchronous RIO count	Number of async RIOs per second in MCU
6			M-JNL Total Number of Journal	The number of journals at MCU
7			M-JNL Average Transfer Rate	The average transfer rate for journals at MCU (KB/sec.)
8			M-JNL Average RIO Response	The remote I/O process time at MCU (milliseconds)
9	RCU	Async copy	R-JNL Asynchronous RIO count	The number of asynchronous remote I/Os per second at the RCU
10			R-JNL Total Number of Journal	The number of journals at the RCU
11			R-JNL Average Transfer Rate	The average transfer rate for journals at RCU (KB/sec.)
12			R-JNL Average RIO Response	The remote I/O process time at RCU (milliseconds)
13	MCU	Journal group	Data Used Rate	Data usage rate for journals at MCU (%)
14			Meta Data Used Rate	Metadata usage rate for journals at MCU (%)
15	RCU	Journal group	Data Used Rate	Data usage rate for journals at RCU (%)
16			Meta Data Used Rate	Metadata usage rate for journals at RCU (%)

Copyright © 2012, Hitachi, Ltd.

(2) HISTORY
This feature enables you to view the history of operations for data volume pairs (Operations performed in the past) using SVP and Web Console. The specifications are as follows.

Table 3.23.7.6.2-3 HISTORY specifications

#	Item	Spec etc.
1	Displayed info	Date and time of the operation, contents of the operation (See #3, Journal group number, Mirror ID, Data volume, Target volume
2	Samplings	524288, or for one week
3	Operation	• Pair create • Pair delete • Pair recovery • Pair split etc.
4	Others	The snapshot function enables you to save operation history to a text file.

Copyright © 2012, Hitachi, Ltd.

3.23.7.6.3 Option

The following table shows UR settings. For pair settings, see 3.23.7.6.1. The following system option and journal group options are supported.

Table 3.23.7.6.3-1 List of options

#	Affected	Setting	Description
1	System	Max number of initial copy VOLs	✓ Number of volumes to which initial copy operations are performed at a time. The setting value is 1-128. ✓ Default is 64.
2	Journal group	Inflow Control	✓ Indicates whether to restrict inflow of update I/Os to the PVOL (whether to delay a response to the hosts). ✓ If you select “Yes”, the inflow control is performed according to the condition of #3.
3		Data Overflow Watch	✓ Indicates the time (in seconds) for monitoring whether metadata and journal data are of journal volume are full. ✓ When either of the areas becomes full, the target group will be suspended due to a failure after this specified time. ✓ The setting value is 0-600 sec.
4		Use of Cache	✓ If you select “USE”, journal data will be stored into the cache and don’t destage journal volumes in the restore journal group as long as the cache memory is not filled. The JNL restore performance improves.

(To be continued)

Copyright © 2012, Hitachi, Ltd.

(Continued from preceding page)

#	Affected	Setting	Description
5	Mirror	Copy pace	✓ Initial copy pace (speed). ✓ The setting value is “Low”, “Medium” or “High”.
6		Speed of Line	✓ JNL copy performance pace (speed). ✓ The setting value is “256Mbps”, “100Mbps” or “50Mbps”.
7		Unit of Path Watch time	✓ Unit of Path Watch time (#8). ✓ The setting value is “minute”, “hour” or “day”.
8		Path Watch time	✓ Time for monitoring a path from the time it is blocked until it is suspended due to a failure. ✓ The setting value is 1 min - 30 days. ✓ You can specify whether to forward the Path Watch time value of the master journal group to the restore journal group. If the Path Watch time value is forwarded from the master journal group to the restore journal group, the two journal groups will have the same Path Watch time value.
9		Forward Path Watch time	✓ If you select “Yes”, the Path Watch time (#8) value of the master journal group will be forwarded to the restore journal group.
10		Delta resync Failure	✓ Specify the processing that takes place when delta resync operation cannot be performed. ✓ If “Copy of all” is specified, the whole data in primary data volume is copied to secondary data volume when delta resync operation cannot be performed.

Copyright © 2012, Hitachi, Ltd.

Table 3.23.7.6.3-2 Link failure monitoring mode

Mode	Feature digest	Feature overview
448	Mode for suspending the pair immediately after UR path is disconnected	Turn this mode ON when you want to suspend the pair due to a failure immediately after a communication failure is detected between UR M-R. This mode runs only when Mode 449 is OFF. ON: In MCU, the pair is suspended due to a failure immediately after the RDJNL from RCU is stopped. In RCU, the pair is suspended due to a failure immediately after the RDJNL fails. OFF: In MCU, the pair is suspended due to a failure when the RDJNL from RCU is stopped for a certain period of time. In RCU, the pair is suspended due to a failure when the RDJNL fails for a certain period of time.
449	Mode for prohibiting UR path watch	Turn this mode ON when you want to prevent communication failures between UR M-R from being detected. ON is default. ON: In MCU, the RDJNL stop check from RCU is prevented. In RCU, the RDJNL failure monitoring is prevented. OFF: Communication failures between M-R are detected.

 Copyright © 2012, Hitachi, Ltd.

3.23.7.7 Maintenance features and procedure

3.23.7.7.1 Maintenance

The following table shows limitations/impact on the maintenance operations regarding the UR operations.

Table 3.23.7.7.1-1 Limitations/restrictions on maintenance

#	Item	Limitation/impact	Note
1	CM replacement/ change of SM capacity setting in MCU	Pair in the initial copy status will suspend due to a failure.	
2	Microprogram exchange in RCU	The pair may be suspended in MCU due to a failure. (Cause: Link failure, journal full)	Alternate paths are set between MCU and RCU. If microprogram is not replaced using these alternate paths at a time, the pair will not be suspended due to a link failure.
3	CHT replacement in MCU/RCU	The pair may be suspended in MCU/RCU due to a failure. (Cause: Link failure, journal full)	Alternate paths are set between MCU and RCU. If microprogram is not replaced using these alternate paths at a time, the pair will not be suspended due to a link failure.

 Copyright © 2012, Hitachi, Ltd.

3.23.7.7.2 PS OFF/ON Process

Before you power off/on (PS OFF/ON) the DKC, we recommend that you suspend all pairs (all JNL groups) by specifying Flush in advance as shown below. When you stop the host I/O and suspend (Flush) the pairs in advance, the contents of PVOL and SVOL match. Consequently, the operation can be continued using the SVOL data in R-DKC even if you fail to power on the M-DKC for some reason. The recommended procedure is as follows.

(1) Stop the host I/O.
(2) Issue the suspend (Flush) request for M-JNL, and change the statuses of both M-JNL and R-JNL to suspend.
(3) Power off the M-DKC.
(4) Power off the R-DKC.
(5) Power on the R-DKC.
(6) Power on the M-DKC.
(7) Resynchronize the pair upon the pair resync request.
(8) Resume the host I/O.

If you do not perform the procedure above, operations are performed as follows. The numbers in the table below show the target of PS OFF/ON, and the order of PS OFF/ON. "-" shows that it is not the target of PS OFF/ON.

Table 3.23.7.7.2-1 PS OFF/ON

	Part and Order of PS OFF/ON				Operation
	M-DKC		R-DKC		
	OFF	ON	OFF	ON	
Case 1	1	2	–	–	[PS OFF/ON only M-DKC] No problem.
Case 2	–	–	1	2	[PS OFF/ON only R-DKC] Pair may be suspended due to a failure in MCU. As a result, R-JNL detects the failure suspension when PS ON is performed, and the pair may be also suspended. Such a failure suspension is caused because the M-JNL is full or the read journal is stopped.
Case 3	1	3	2	4	[PS OFF/ON both M-DKC and R-DKC] PS ON the M-DKC first. If it takes time until R-DKC PS ON (4) after M-DKC PS ON (3), the same phenomenon as Case 2 will occur.
Case 4	1	4	2	3	[PS OFF/ON both M-DKC and R-DKC] The procedure is the recommended one, except that host I/O stop and pair suspension are excluded. When PS ON is performed, the pair status before PS OFF is maintained.
Case 5	2	3	1	4	[PS OFF/ON both M-DKC and R-DKC] PS OFF R-DKC first. If it takes time until M-DKC PS OFF (2) after R-DKC PS OFF (1), the same phenomenon as Case 2 will occur. If it takes time until R-DKC PS ON (4) after M-DKC PS ON (3), the same phenomenon as Case 2 will occur.
Case 6	2	4	1	3	[PS OFF/ON both M-DKC and R-DKC] If it takes time until M-DKC PS OFF (2) after R-DKC PS OFF (1), the same phenomenon as Case 2 will occur.

 Copyright © 2012, Hitachi, Ltd.

3.23.7.7.3 Power failure

When a power failure occurs, UR pairs will be in the following status. The status change is the same in both the memory backup mode and the destage mode.

Table 3.23.7.7.3-1 Power failures

#	Item	Pair in PS ON	Recovery
1	SM/CM non-volatile	Suspended due to a failure	Resynchronize to the target group
2	SM/CM volatile	Journal group information and pair information will be lost.	Register journal group and pair

Copyright © 2012, Hitachi, Ltd.

3.23.7.8 Cautions on software

3.23.7.8.1 Error recovery

The following subsections describes failure detection, error report, and recovery in UR. When a failure occurs in UR, Group or Volume (according to the error level setting) will be suspended.

(1) Error report
The following table shows the error report in the case of a failure.

Table 3.23.7.8.1-1 Error report

Item	Type	SSB (F/M)	SIM
Path	Link failure	–	0x2180-XX
Pair	Failure suspend	F/M=0xFB Not reported to host	Serious SIM Not reported to host

(2) Failure detection and recovery
Failure detection and recovery in MCU are described in the following table.

Table 3.23.7.8.1-2 Failure detection and recovery

Part	Description	Error	Recovery
Path failure	Link failure of MCU->RCU	SIM=0x2180	Recover the link
	Link failure of MCU<-RCU	When Read JNL is stopped for a certain period of time, the target Group is suspended, and F/M=0xFB, SIM is generated. (see Table 3.23.7.6.3-1 and Table 3.23.7.6.3-2.)	Recover the link Resync after the failure recovered
Data volume failure	PDEV failure	Target Volume/Group is suspended F/M=0xFB, SIM Update I/O is managed as differential data	Resync after the failure recovered
Journal failure	PDEV failure	Target Group is suspended F/M=0xFB, SIM Update I/O is managed as differential data	Resync after the failure recovered
Journal full	Metadata area or journal data area of JNL volume is not sufficient for a certain period of time	Target Group is suspended F/M=0xFB, SIM Update I/O is managed as differential data	Resync after the failure recovered

 Copyright © 2012, Hitachi, Ltd.

(3) Failure detection in RCU

Failure detection and recovery in RCU are described in the following table.

Table 3.23.7.8.1-3 Failure detection in RCU

Part	Description	Error	Recovery
Path failure	Link failure of MCU<-RCU	SIM=0x2180When Read JNL is stopped for a certain period of time, the target Group is suspended, and F/M=0xFB, SIM is generated. (see Table 3.23.7.6.3-1 and Table 3.23.7.6.3-2.)	Recover the link Resync after the failure recovered
Data volume failure	PDEV failure	Target Volume/Group is suspended F/M=0xFB, SIM	Resync after the failure recovered
Journal failure	PDEV failure	Target Group is suspended F/M=0xFB, SIM	Resync after the failure recovered

Copyright © 2012, Hitachi, Ltd.

3.23.7.9 Disaster Recovery of UR

3.23.7.9.1 Preparing for Disaster Recovery

The type of disaster and the status of the Universal Replicator volume pairs determines the best approach for disaster recovery. Unfortunately, some disasters are not so “orderly” and involve intermittent or gradual failures occurring over a longer period of time. The major steps in preparing for disaster recovery are:

1. Identify the journal groups and volumes that contain important files and data for disaster recovery.
2. Install the Storage Navigator computer and Universal Replicator hardware and software, and establish Universal Replicator operations for the journal groups and volumes identified in step 1.
3. Establish file and database recovery procedures. These procedures should already be established for recovering volumes that become inaccessible due to some failure.
4. Install and configure host failover software between the primary and secondary sites.

3.23.7.9.2 File and Database Recovery Procedures

When the primary or secondary storage system suspends a Universal Replicator pair due to a disaster, the secondary volume may contain in process data. A data set could be open, or transactions may not have completed. Therefore, you need to establish file recovery procedures. These procedures should be the same as those used for recovering volume that becomes inaccessible due to control unit failure.

Universal Replicator does not provide any procedure for detecting and retrieving lost updates. To detect and recreate lost updates, you must check other current information (for example, database log file) that was active at the primary site when the disaster occurred. Since this detection/retrieval process can take a while, your disaster recovery scenario should be designed so that detection/retrieval of lost updates is performed after the application has been started at the secondary site.

You should prepare for file and database recovery by using files for file recovery (for example, database log files that have been verified as current). See “15.4 Pinned Track Recovery Procedure for UR” (TRBL15-120) for information on recovering a pinned track on a Universal Replicator volume.

3.23.7.9.3 Switching Operations to the Secondary Site

Executing the **`horctakeover`** command of CCI (Command Control Interface) on Universal Replicator pairs enables your business tasks to be taken over by the secondary site, using secondary volumes in the secondary site. If a disaster or failure occurs at the primary site, the first disaster recovery activity is to switch your operations to the secondary site.

- The **`horctakeover`** command checks the pair status of secondary volumes and splits journal groups that will become a Universal Replicator pair. This ensures consistency of secondary volumes and makes the secondary volumes usable.
- If possible, the **`horctakeover`** command attempts to restore pairs to reverse primary volumes and secondary volumes.
- If the **`horctakeover`** command executes successfully, your business tasks are taken over to (or your business application is started at) the secondary site, using secondary volumes. For detailed information about how to use CCI, see the Command Control Interface User and Reference Guide.

Copyright © 2012, Hitachi, Ltd.

3.23.7.9.4 Transferring Operations Back to the Primary Site

Once the disaster recovery procedure is finished and your business applications are running at the secondary site, the next activity is to restore the primary site and make arrangements for copying data from the secondary site back to the primary site. The following procedure explains how to use CCI to copy data from the secondary site to the primary site:

1. Restore the primary storage system and remote copy connections, bring up the host server(s) at the primary site, and make sure that all Universal Replicator components are fully operational.
2. At the primary site, locate primary volumes whose pair status is COPY or PAIR, and then locate corresponding secondary volumes whose pair status is SSWS. If such volume pairs are found, issue a request for splitting the pairs to the primary volumes.
3. At the primary site, locate primary volumes whose pair status is not SMPL, and then locate corresponding secondary volumes whose pair status is SMPL. If such volume pairs are found, issue a request for releasing the pairs to the primary volumes.
4. At the primary site, locate volume pairs whose pair status is SMPL, and then issue a request for releasing the pairs to the secondary volumes.
5. Execute the **`pairresync -swaps`** command on secondary volumes whose pair status is SSWS (**`pairresync`** is the CCI command for resynchronizing pair and **`-swaps`** is a swap option). This reverses primary volumes and secondary volumes to resynchronize pairs.
6. Create pairs, specifying secondary volumes whose pair status is SMPL as primary volumes. This creates pairs in which primary volumes and secondary volumes are reversed.
7. Verify that pair status of all secondary volumes (which were originally primary volumes) changes from COPY to PAIR. If the pair status is changed to PAIR, initial copy operations are finished and consistency is maintained.

The above procedure enables copying of data from the secondary site to the primary site. Data in the secondary site is reflected on the primary site.

For detailed information about how to use CCI, see the Command Control Interface User and Reference Guide.

For detailed information about status of volumes, see Pair Status.

3.23.7.9.5 Resuming Normal Operations at the Primary Site

Once the Universal Replicator volume pairs have been established in the reverse direction, you are ready to resume normal operations at the primary site. The following procedure explains how to resume normal operations at the primary site by using CCI. Remember that the Universal Replicator terminology is now reversed: the original primary volumes are now secondary volumes, and the original secondary volumes are now primary volumes.

1. At the primary and secondary sites, make sure that all Universal Replicator components are fully operational and are free from failures.
2. Make sure that pair status of primary and secondary volumes in all Universal Replicator pairs is “PAIR”. This indicates that the Universal Replicator initial copy operations are complete and consistency is maintained.
3. Stop the applications at the secondary site.
4. Issue a request for splitting pairs to master journal groups (which were originally restore journal groups); use the Flush option when splitting pairs. If an error occurs when splitting pairs, remove the error cause and go back to step 1 after resuming your business task at the secondary site.
5. If no error occurs in step 4, wait until suspension finishes. After suspension finishes, check whether there is a secondary volume (which is originally a primary volume) whose pair status is other than PSUS. If such a pair exists, remove the error cause and go back to step 1 after resuming your business task at the secondary site.
6. If there is no secondary volume (which is originally a primary volume) whose pair status is other than PSUS, data in primary volumes are the same as data in secondary volumes, and the secondary volume (which are originally primary volumes) are usable. Resume applications at the primary site.
7. Execute the **`pairresync -swaps`** command (**`pairresync`** is the CCI command for resynchronizing pair and **`-swaps`** is a swap option). This reverses primary volumes and secondary volumes to resynchronize pairs.

For detailed information about how to use CCI, see the Command Control Interface User and Reference Guide.

For detailed information about status of volumes, see “Hitachi Universal Replicator User Guide”.

Copyright © 2012, Hitachi, Ltd.

(1) Means for setting the paired LU:
The following three means are provided:
- RAID Manager/UR
- Web Console
- Storage Navigator

Not only the pairing but also a series of pair state changes are possible by using these three means. However, the user can use two means only: the command instruction from the RAID Manager/UR and the instruction from the Web Console.

(2) LU size capable of the duplicated writing:
If two LUSE volumes are paired with UR, a LUSE P-VOL must be paired with S-VOL of the same size and the same structure. For example, if a LUSE P-VOL is connected with the volumes of 1GB, 2GB, and 3GB in this order, you must specify the LUSE volume which has exactly the same size and the same connection order as the S-VOL.

(3) Copy Mode:
UR: The copy operation and the host I/O can be performed asynchronously, but it must to ensure the update sequence consistency of Write progress across multiple primary volumes (The data written late cannot be copied earlier.). In addition, when a failure occurs, the function (for multiple pairs) having multiple pairs blocked while keeping the update sequence consistency is available. In this way, the group composed of pairs, which are the control objects, is called Consistency Group.

(4) Means for setting the Journal Volume:
The following two means are provided.
- Web Console
- Storage Navigator

Not only the setting Journal volume but also Journal Group options are possible to be changed by using these two means. However, the user can use only instruction from the Web Console.

Copyright © 2012, Hitachi, Ltd.

◆ Restrictions:

(1) Command device:

① The UR provides users with a command to enable a state change and status display of the UR pair from the server.

② Assign a special LUN called a command device so that the DW700 can receive this pair state change and pair status display commands.

③ Users cannot use the command device. The capacity allocated to command devices in a storage system is not available for users. The minimum capacity of command device is 47MB.

④ Use Web Console to specify the command device.

(2) Flashing updated data in the server:

When the UR is used as a DataPlex function, split the primary/secondary paired VOL. A Sync command or the like must be issued before splitting it and a file system buffer must be flashed when acquiring a backup from the secondary VOL. Thus, the latest backup can be acquired.

(3) S-VOL (secondary VOL) access:

① An RD access to the secondary VOL is permitted to accept the RD command issued to the disk label when the secondary server is started.

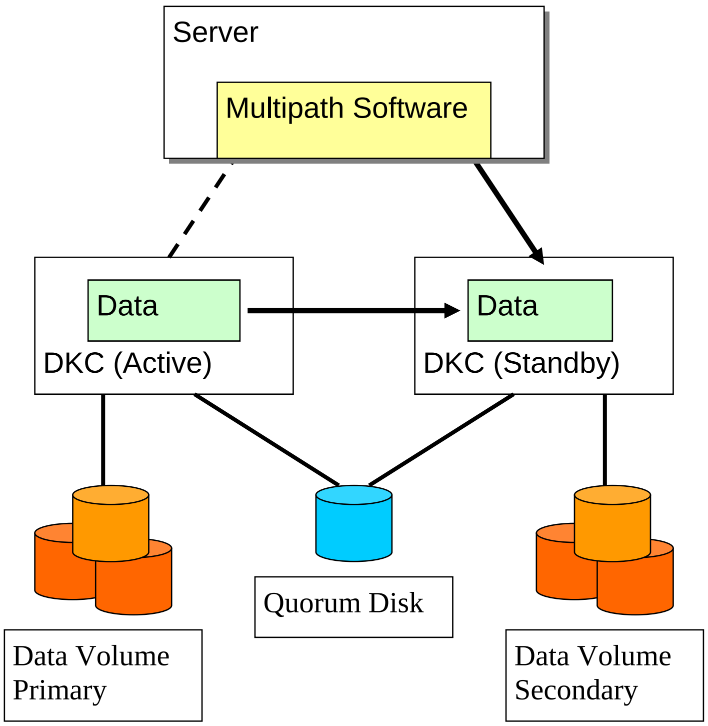
② In order to support the DataPlex function, write access to the secondary VOL is permitted on condition that the pair is being suspended.
Using the RAID Manager/UR or SVP, you can indicate the permission of write operation to S-VOL. After this indication, if the server performs any write operation to S-VOL, in Pair Resync (Resume) operation all tracks on P-VOL will be copied to S-VOL. If using SVP, the permission of write operation to S-VOL is executed by setting “S-VOL write Enable” on Suspend Pair display in the indication of S-VOL Suspend on MCU.
Also, you can confirm using RAID Manager/UR or SVP whether “S-VOL write Enable” on S-VOL is permitted or not.

Copyright © 2012, 2013, Hitachi, Ltd.

3.23.7.10 HAM (High Availability Manager) Overview

The High Availability Manager function can duplicate data (volumes) under the control of the storage system by dire directly connecting the two DKCs as much as TrueCopy.
When the trouble occurs in active DKC, it is necessary to run your applications to the TrueCopy function. However, the HAM can operate failover by the automatic operation.
There is Fibre channel interface of connection form between CUs.

Outline of High Availability Manager



 Copyright © 2013, Hitachi, Ltd.

3.23.7.11 Basic HAM Specifications

Basic HAM specifications are shown below.

No.	Item	Description	Remarks
1	Host interface on open system side	Fibre Channel	Conforming to the TrueCopy
2	Supporting OS platform	Complied with the supported platforms of HDLM.	
3	Supporting Multipath Software	HDLM	
4	Connection between the CUs	Fibre Channel	Conforming to the TrueCopy
5	Number of LUs capable of the HAM Volume	Maximum 16,384 pairs	Conforming to the TrueCopy
6	LU type for setting the HAM volume	Normal VOL	Maximum 3TB
		DP-VOL	Maximum 4TB
		LUSE	
7	Means for setting the paired LU	Storage Navigator	
		Web Console	
		Raid Manager (*1)	
8	Fence level	Never	
9	Number of LUs capable of the Quorum Disk	Maximum 128	
10	LU type for setting the Quorum Disk	Only External Volume	
11	Means for setting the Quorum Disk	Storage Navigator	
		Web Console	

*1: When you use Raid Manager, the HAM volumes are seen as the TrueCopy volumes. Operation for HAM volumes is available in Raid Manager for only Suspend, SwapSuspend, Resync, or Reverse-RESYNC.

◆ Means for setting the paired LU:

The following three means are provided:

- RAID Manager
- Web Console
- Storage Navigator

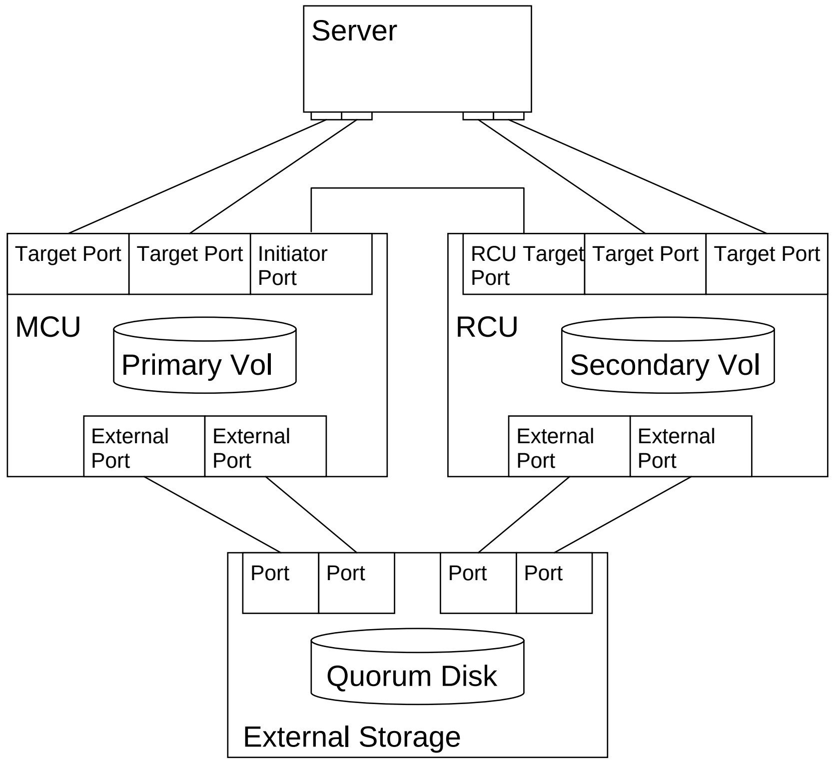
Both creating a pair and changing a series of pair statuses are available by using Web Console or Storage Navigator.

 Copyright © 2013, Hitachi, Ltd.

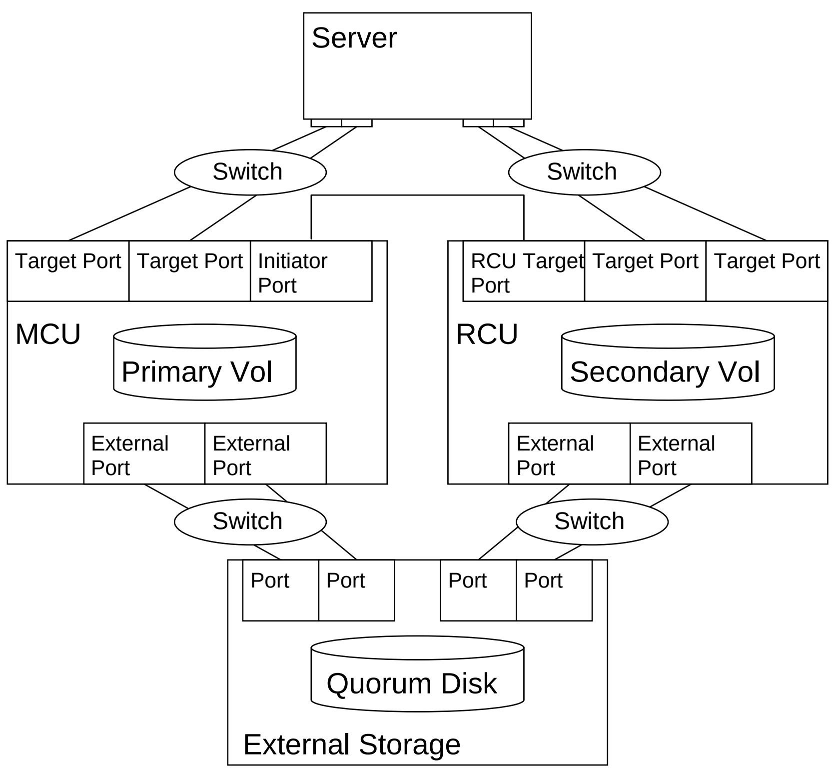
◆ Connected block diagram

Connected composition example is shown in the following.

(a) Direct

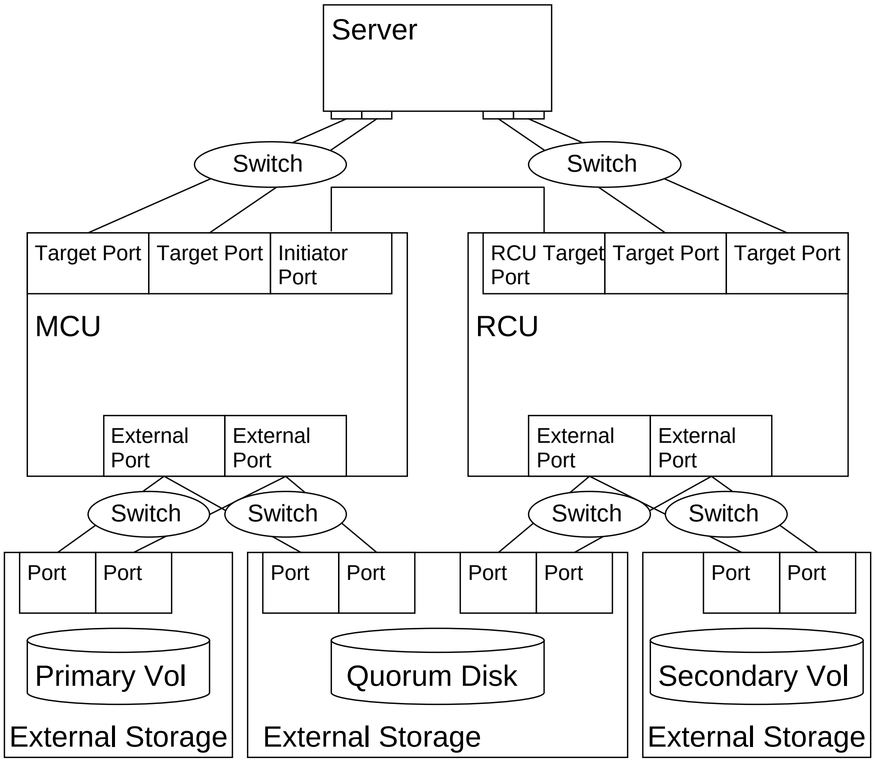


(b) Between switch



Copyright © 2013, Hitachi, Ltd.

(c) When you use the external storage for primary vol or secondary vol.



Copyright © 2012, Hitachi, Ltd.

3.23.8 LUN installation

3.23.8.1 Overview

LUN installation feature makes it enable to add LUNs to DW700 Fibre ports while I/Os are still running.
Some host operations are required before the added volumes are recognized and become usable from the host operating systems.
MCU port (Initiator port)/External of TrueCopy function does not support LUN installation.

3.23.8.2 Specifications

(1) General
- (a) LUN installation feature supports Fibre interface.
- (b) LUN installation is supported.
- (c) LUN installation can be executed by SVP or by Web Console.
- (d) Some operating systems require reboot operation to recognize the newly added volumes.
- (e) When new LDEVs should be installed for LUN installation, install the LDEVs by SVP first. Then add LUNs by LUN installation from SVP or Web Console.
- (f) MCU (Initiator port)/External port of TrueCopy function does not support LUN installation.

(2) Platform support
Host Platforms supported for LUN installation are shown in Table 3.23.8.2-1.

Table 3.23.8.2-1 Platform support level

Support level	FIBRE
(A) LUN installation and LUN recognition.	Solaris, HP-UX, AIX, Windows Server 2003 Windows Server 2008
(B) LUN installation only. Reboot is required before new LUNs are recognized.	Linux
(C) LUN installation is not supported. Host must be shutdown before installing LUNs and then must be rebooted.	—

3.23.8.3 Operations

(1) Operations
Step 1: Execute LUN installation from SVP or from JAVA = “Web Console”.
Step 2: Check whether or not the initiator platform of the Fibre port supports LUN recognition with Table 3.23.8.3-1.
Support (A) -> Execute LUN recognition procedures in Table 3.23.8.3-1
Not support (B) -> Reboot host and execute normal install procedure.

(2) Host operations
Host operations for LUN recognition are shown in Table 3.23.8.3-1.

Table 3.23.8.3-1 LUN recognition procedures outline for each platform

Platform	LUN recognition procedures
HP-UX	(1) **ioscan** (check device added after IPL) (2) **insf** (create device files)
Solaris	(1) **/usr/sbin/drvconfig** (2) **/usr/sbin/devlinks** (3) **/usr/sbin/disks** (4) **/usr/ucb/ucblinks**
AIX	(1) **Devices-Install/Configure Devices Added After IPL** By SMIT
Windows Server 2003 Windows Server 2008	Automatically detected

3.23.9 LUN de-installation

3.23.9.1 Overview

LUN de-installation feature makes it enable to delete LUNs to DW700 Fibre ports while I/Os are still running.
MCU (Initiator port)/External port of TrueCopy function does not support Online LUN de-installation.

3.23.9.2 Specifications

(1) General
- (a) LUN de-installation feature supports Fibre interface.
- (b) LUN de-installation can be used only for the ports on which LUNs are already existing.
- (c) LUN de-installation can be executed by SVP or by “Web Console”.
- (d) When LUNs should be de-installed for LUN de-installation, stop Host I/O of concerned LUNs.
- (e) If necessary , execute backup of concerned LUNs.
- (f) De-install concerned LUNs from HOST.
- (g) In case of AIX, release the reserve of concerned LUNs.
- (h) In case of HP-UX do not delete LUN=0 under existing target ID.
- (i) MCU (Initiator port)/External port of TrueCopy function does not support Online LUN de-installation.

NOTE: If LUN de-installation is done without stopping Host I/O, or releasing the reserve, it would fail. Then stop HOST I/O or release the reserve of concerned LUNs and try again. If LUN de-installation would fail after stopping Host I/O or releasing the reserve, there is a possibility that the health check command from HOST is issued. At that time, wait about three minutes and try again.

(2) Platform support
Host platforms supported for LUN de-installation are shown in Table 3.23.9.2-1.

Table 3.23.9.2-1 Support platform

Platform	OS	Fibre
HP	HP-UX	○
SUN	Solaris	○
RS/6000	AIX	○
PC	Windows Server 2003 Windows Server 2008	○

(example) ○: support, ×: not support

3.23.9.3 Operations

(1) Operations

Step 1: Confirm whether or not the initiator platform of the FIBRE port supports LUN de-installation with Table 3.23.9.2-1.
Support :Go to Step 2.
Not support :Go to Step 3.

Step 2: If HOST MODE of the port is not 00 or 04 or 07 use, go to Step 4.

Step 3: Stop Host I/O of concerned LUNs.

Step 4: If necessary, execute backup of concerned LUNs.

Step 5: De-install concerned LUNs form HOST.

Step 6: In case AIX, release the reserve of concerned LUNs.
If not, go to Step 7.

Step:7 Execute LUN de-installation from SVP or from Remote “Web Console”.

(2) Host operations

Host operations for LUN de-installation procedures are shown in Table 3.23.9.3-1.

Table 3.23.9.3-1 LUN de-installation procedures outline for each platform

Platform	LUN de-installation procedures	
HP-UX	mount point:/01, volume group name:vg01	
	(1) **umount /01**	(umount)
	(2) **vgchange -a n vg01**	(deactive volume groups)
	(3) **vgexport /dev/vg01**	(export volume groups)
Solaris	mount point:/01	
	(1) **umount /01**	(unmout)
AIX	mount point:/01, volume group name:vg01, device file name:hdisk1	
	(1) **umount /01**	(umount)
	(2) **rmfs -r” /01**	(delete file systems)
	(3) **varyoffvg vg01**	(vary off)
	(4) **exportvg vg01**	(export volume groups)
	(5) **rmdev -I ‘hdisk1’ ‘-d’**	(delete devime files)

Copyright © 2012, Hitachi, Ltd.

3.23.10 Prioritized Port Control (PPC)

3.23.10.1 Overview

The Prioritized Port Control (PPC) feature allows you to use the DKC for both production and development. The assumed system configuration for using the Prioritized Port Control option consists of a single DKC that is connected to multiple production servers and development servers. Using the Prioritized Port Control function under this system configuration allows you to optimize the performance of the development servers without adversely affecting the performance of the production servers.
MCU port (Initiator port) of Fibre Remote Copy function does not support Prioritized Port Control (PPC).

The Prioritized Port Control option has two different control targets: fibre port and open-systems host's World Wide Name (WWN). The fibre ports used on production servers are called prioritized ports, and the fibre ports used on development servers are called non-prioritized ports. Similarly, the WWNs used on production servers are called prioritized WWNs, and the WWNs used on development servers are called non-prioritized WWNs.

NOTE: The Prioritized Port Control option cannot be used simultaneously for both the ports and WWNs for the same DKC. Up to 176 ports or 2048 WWNs can be controlled for each DKC.

The Prioritized Port Control option monitors I/O rate and transfer rate of the fibre ports or WWNs. The monitored data (I/O rate and transfer rate) is called the performance data, and it can be displayed in graphs. You can use the performance data to estimate the threshold and upper limit for the ports or WWNs, and optimize the total performance of the DKC.

- Prioritized Ports and WWNs
  The fibre ports or WWNs used on production servers are called prioritized ports or prioritized WWNs, respectively. Prioritized ports or WWNs can have threshold control set, but are not subject to upper limit control. Threshold control allows the maximum workload of the development server to be set according to the workload of the production server, rather than at an absolute level. To do this, the user specifies whether the current workload of the production server is high or low, so that the value of the threshold control is indexed accordingly.

- Non-Prioritized Ports and WWNs
  The fibre ports or WWNs used on development servers are called non-prioritized ports or prioritized WWNs, respectively. Non-prioritized ports or WWNs are subject to upper limit control, but not threshold control. Upper limit control makes it possible to set the I/O of the non-prioritized port or WWN within a range that does not affect the performance of the prioritized port or WWN.

3.23.10.2 Overview of Monitoring

- Monitoring Function
  Monitoring allows you to collect performance data, so that you can set optimum upper limit and threshold controls. When monitoring the ports, you can collect data on the maximum, minimum and average performance, and select either per port, all prioritized ports, or all non-prioritized ports. When monitoring the WWNs, you can collect data on the average performance only, and select either per WWN, all prioritized WWNs, or all non-prioritized WWNs.
  The performance data can be displayed in graph format either in the real time mode or offline mode. The real time mode displays the performance data of the currently active ports or WWNs. The data is refreshed in every time that you specified between 1 and 15 minutes by minutes, and you can view the varying data in real time. The offline mode displays the stored performance data. Statistics are collected at a user-specified interval between 1 and 15 minutes, and stored between 1 and 15 days.

- Monitoring and Graph Display Mode
  When you activate the Prioritized Port Control option, the Select Mode panel where you can select either Port Real Time Mode, Port Offline Mode, WWN Real Time Mode, or WWN Offline Mode opens. When you select one of the modes, monitoring starts automatically and continues unless you stop monitoring. However, data can be stored for up to 15 days. To stop the monitoring function, exit the Prioritized Port Control option, and when a message asking if you want to stop monitoring is displayed, select the Yes button.
  - The Port/WWN Real Time Mode is recommended if you want to monitor the port or WWN performance for a specific period of time (within 24 hours) of a day to check the performance in real time.
  - The Port/WWN Offline Mode is recommended if you want to collect certain amount of the port or WWN performance data (maximum of one week), and check the performance in non-real time.

To determine a preliminary upper limit and threshold, run the development server by using the performance data collected from the production server that was run beforehand and check the changes of performance of a prioritized port. If the performance of the prioritized port does not change, set a value by increasing an upper limit of the non-prioritized port. After that, recollect and analyze the performance data. Repeat these steps to determine the optimized upper limit and threshold. (See Fig. 3.23.10.3-1.)

Copyright © 2012, Hitachi, Ltd.

3.23.10.3 Procedure (Flow) of Prioritized Port Control

To perform the prioritized port control, determine the upper limit to the non-prioritized port by checking that the performance monitoring function does not affect production. Fig. 3.23.10.3-1 shows the procedures for prioritized port control.

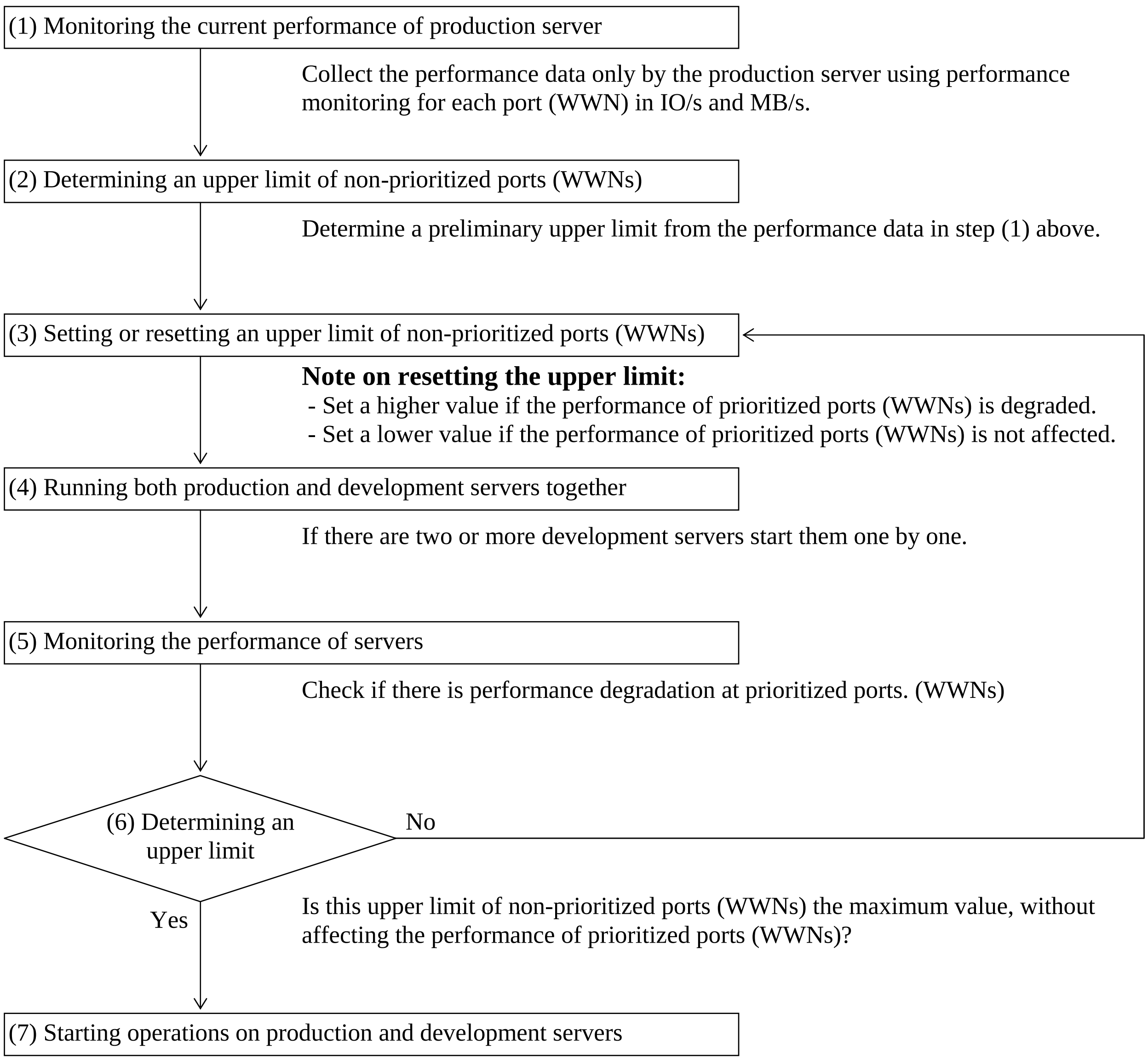


Fig. 3.23.10.3-1 Flow of Prioritized Ports Control

Copyright © 2013, Hitachi, Ltd.

3.24 Notes on maintenance during LDEV Format/drive copy operations

This section describes whether maintenance operations can be performed when Dynamic Sparing, Correction Copy, Copy Back, or LDEV Format is running or when data copying to a spare disk is complete.
If Correction Copy runs due to a drive failure or Dynamic Sparing runs due to preventive maintenance on large-capacity disk drives or flash drives, it may take long to copy data. In the case of low-speed LDEV Format performed due to volume installation, it may take time depending on the I/O frequency because host I/Os are prioritized. In such a case, it is recommended to perform operations, such as replacement, installation, and removal, after Dynamic Sparing, LDEV Format etc. is completed, based on the basic maintenance policy, but the following maintenance operations are available.

Copyright © 2013, Hitachi, Ltd.

Maintenance operation		Storage system status				
		Dynamic Sparing	Correction Copy	Copy Back	Copied to spare disk	LDEV Format
Replacement	MAIN	Impossible	Impossible	Impossible	Possible	Impossible
	MPB	Impossible	Impossible	Impossible	Possible	Impossible
	CHB	Impossible	Impossible	Impossible	Possible	Impossible
	Power supply	Possible	Possible	Possible	Possible	Possible
	SVP	Possible	Possible	Possible	Possible	Possible
	ENC	Impossible	Impossible	Impossible	Possible	Impossible
	DKB	Impossible	Impossible	Impossible	Possible	Impossible
	PDEV	Possible (*1)	Possible (*1)	Possible (*1)	Possible	Possible (*4) (*6)
Installation/ Removal	CACHE/SM	Impossible	Impossible	Impossible	Possible (*6)	Impossible
	MPB	Impossible	Impossible	Impossible	Possible (*6)	Impossible
	CHB	Impossible	Impossible	Impossible	Possible (*6)	Impossible
	SVP	Possible	Possible	Possible	Possible	Possible
	DKB	Impossible	Impossible	Impossible	Possible (*6)	Impossible
	PDEV	Impossible	Impossible	Impossible	Possible (*2) (*6)	Impossible
Micro-program exchange	Online	Possible (*3) (*6)	Possible (*3) (*6)	Possible (*3) (*6)	Possible (*1) (*7)	Impossible
	Offline	Impossible	Impossible	Impossible	Possible	Impossible
	SVP only	Possible	Possible	Possible	Possible	Possible
LDEV maintenance	Blockade	Impossible	Impossible	Impossible	Possible	Impossible
	Restore	Impossible	Impossible	Impossible	Possible	Impossible
	Format	Possible (*5) (*6)	Possible (*5) (*6)	Possible (*5) (*6)	Possible	Impossible
	Verify	Impossible	Impossible	Impossible	Possible	Impossible

*1: It is prevented with a message, but you can perform it by entering the password.

*2: It is impossible to remove a RAID group in which data is migrated to a spare disk and the spare disk.

*3: Micro-program exchange can be performed if HDD micro-program exchange is not included.

*4: It is impossible when high-speed LDEV Format is running. When low-speed LDEV Format is running, it is possible to replace PDEV in a RAID group in which LDEV Format is not running.

*5: It is possible to perform LDEV Format for LDEV defined in a RAID group in which Dynamic Sparing, Correction Copy, or Copy Back is not running.

*6: It is available when the micro-program version is 73-02-xx-xx/xx or later.

*7: Micro-program exchange can be performed if HDD or Expander or Expander (DBF) micro-program exchange is not included.

Copyright © 2013, Hitachi, Ltd.

3.25 Encryption License Key

3.25.1 Overview of encryption

In DW700 storage system, you are able to encrypt data stored in volumes in the storage system by using Encryption License Key.
By encrypting data, you can prevent data from being leaked when you replace the storage system or hard disks in the storage system or when these are stolen.

3.25.2 Specifications of encryption

Table 3.25.2-1 shows the specifications of encryption with Encryption License Key.

Table 3.25.2-1 Specifications of encryption with Encryption License Key

Item		Specifications
Hardware spec	Encryption algorithm	AES 256bit
Volume to encrypt	Volume type	Open volumes
	Emulation type	OPEN-V
Encryption key management	Unit of creating encryption key	PDEV
	Number of encryption keys	1,536
	Unit of setting encryption	RAID group
Converting non-encrypted data/ encrypted data	Encryption of existing data	Convert non-encrypted data/encrypted data for RAID group in which encryption is set by using the existing function (Volume Migration, ShadowImage etc.).

Copyright © 2013, Hitachi, Ltd.

3.25.3 Notes on using Encryption License Key

Please note the following when using Encryption License Key.

Table 3.25.3-1 Notes on Encryption License Key

#	Item	Content
1	Volumes to encrypt	Only internal volumes in the disk subsystem can be encrypted with Encryption License Key. External volumes cannot be encrypted.
2	LDEV format	You cannot perform high-speed format (drive format) for disks under RAID group in which encryption is set. Time required for LDEV format performed for the RAID group in which encryption is set depends on the number of RAID groups.
3	Encryption/ non-encryption status of volumes	When encrypting user data, set encryption for all volumes in which the data is stored in order to prevent the data from being leaked. Example 1: In the case a copy function is used, and when encryption is set for P-VOL, set it also for S-VOL. When encryption is set for P-VOL, and non-encryption is set for S-VOL (or vice versa), you cannot prevent data on the non-encrypted volume from being leaked. Example 2: To encrypt data stored in LUSE volumes or pool, create the volume by using only encrypted LDEVs. If you create the volume by using both encrypted LDEVs and non-encrypted LDEVs, you cannot prevent the data from being leaked.
4	Convert non-encrypted data/ encrypted data	The specifications of converting non-encrypted data/encrypted data comply with the specifications of the copy function (Volume Migration, ShadowImage etc.) used for conversion.
5		In the case of Volume Migration, you cannot perform such an auto migration that would switch encryption/non-encryption status.
6		To convert non-encrypted LUSE volumes into encrypted ones with auto migration of Volume Migration, you need to convert each LDEV that constitutes LUSE. Therefore while LUSE volume migration is performed, LUSE contains both encrypted volumes and non-encrypted volumes. To prevent data from being leaked, complete conversion of all volumes that constitute LUSE when converting encrypted data into encrypted data.
7	Switch encryption setting	When you switch the encryption setting of RAID group, you need to perform LDEV format again. To switch the encryption setting, back up data as necessary.
8	[Protect the Key Encryption Key at the Key Management Server] Check Box	When you select [Protect the Key Encryption Key at the Key Management Server] Check Box, the KEK is stored in the key management server. Registration of 2 key management servers is required for this operation. When you perform PS-ON for the DKC, the SVP gets the key from the key management server. Therefore, the communication between the SVP and the key management server should be available. Before you perform PS-ON for the DKC, make sure that the communication between the SVP and the key management server is available.

Copyright © 2013, Hitachi, Ltd.

3.25.4 Creation of encryption key

An encryption key is used for data encryption and decryption. 1,536 encryption keys can be created in a storage system.
Only customer security administrators are able to create encryption keys.

In the following cases, however, creation of encryption key is inhibited to avoid data corruption.

- Due to a failure in the disk subsystem, the disk subsystem does not have any encryption key but it has a RAID group in which encryption is set.
  In this case, restore the backed up encryption key.

3.25.5 Backup of encryption key

There are two types of encryption key backup; backup within the subsystem and backup outside the subsystem. In the case of backup within the subsystem, the encryption key is backed up in the flash memory in the disk subsystem. In the case of backup outside the subsystem, it is backed up in the machine in which Storage Navigator is running (Storage Navigator PC).

- Backup within the subsystem
  Encryption key created on SM is backed up in the flash memory in the disk subsystem.
  Encryption key is automatically backed up within the subsystem at the time it is created, deletion, a state are changed.

- Backup outside the subsystem
  Encryption key created on SM is backed up in the Storage Navigator PC. Backup outside the subsystem is performed when requested by the user from Storage Navigator.

Copyright © 2013, Hitachi, Ltd.

3.25.6 Restoration of encryption key

There are two types of encryption key restoration; restoration from backup within the subsystem and restoration from backup outside the subsystem.

- Restoration from backup within the subsystem
  When the encryption key on SM cannot be used, the encryption key backed up within the subsystem is restored.
  Restoration from backup within the subsystem is automatically performed in the subsystem.

- Restoration from backup outside the subsystem
  When encryption keys including the encryption key backed up within the subsystem cannot be used in the disk subsystem, the encryption key backed up outside the subsystem is restored.
  Restoration from backup outside the subsystem is performed when requested by the security administrator from Storage Navigator.

3.25.7 Setting and releasing encryption

You can set and release encryption by specifying a RAID group. Set and release encryption in the Parity Group list window in Web Console. To set encryption in a RAID group, you have to create the same number of encryption keys which are not allocated or more as the number of HDDs which configure the RAID group.

NOTE:
- Encryption can be set and released only when all volumes that belong to the RAID group are blocked, or when there is no volume in the RAID group.
  When the RAID group contains at least one volume that is not blocked, you cannot set and release encryption.
- When you switch the encryption setting, you need to perform LDEV format again. Therefore set encryption before formatting the entire RAID group when installing RAID groups etc.

Copyright © 2013, Hitachi, Ltd.

3.25.8 Encryption format

To format a RAID group in which encryption is set, format the entire disk area by writing encrypted 0 data in the entire disk area. This is called Encryption format.
When encryption is set for a RAID group, encryption format is needed before user data is written.
When encryption is released, normal format is needed before user data is written.

NOTE: Encryption format can be performed only when all volumes in the RAID group can be formatted. When at least one volume cannot be formatted, encryption format cannot be performed.

3.25.9 Converting non-encrypted data/encrypted data

To encrypt existing data, create a RAID group in which encryption is set in advance and use a copy Program Product, such as Volume Migration, and ShadowImage, to convert data. Data conversion is performed per LDEV.
The specifications of converting non-encrypted data/encrypted data comply with the specifications of the copy function (Volume Migration, ShadowImage etc.) used for conversion.

3.25.10 Deleting encryption keys

Only customer security administrators are able to delete encryption keys. You cannot delete the encryption key which is allocated to HDDs or DKBs. You can delete the encryption key whose attribute is Free key.

3.25.11 Reference of encryption setting

You can check the encryption setting (Encryption: Enable/Disable) per RAID group from [Install]-[Refer Configuration] ‘Device Configuration’ screen of SVP or Parity Groups screen of Storage Navigator.

Copyright © 2012, Hitachi, Ltd.

4. Power-on Sequences

4.1 IMPL Sequence

The IMPL sequence, which is executed when power is turned on, is comprises of the following four modules:

(1) BIOS
The BIOS starts other MP cores after a ROM boot. Subsequently, the BIOS expands the OS loader from the flash memory into the local memory and OS loader is executed.

(2) OS loader
The OS loader performs the minimum necessary amount of initializations, tests the hardware resources, then loads the Real Time OS modules into the local memory and the Real Time OS is executed.

(3) Real Time OS modules
Real Time OS is a root task that initializes the tables in the local memory that are used for intertask communications. Real Time OS also initializes network environment, and create the DKC task.

(4) DKC task
When the DKC task is created, it executes initialization routines. Initialization routines initialize the most part of the environment that the DKC task uses. When the environment is established so that the DKC task can start scanning, the DKC task notifies the SVP of a power event log. Subsequently, the DKC task turns on the power for the physical drives and, when the logical drives become ready, The DKC task notifies the host processor of an NRTR.

The control flow of IMPL processing is shown in Fig. 4.1-1.

Copyright © 2012, Hitachi, Ltd.

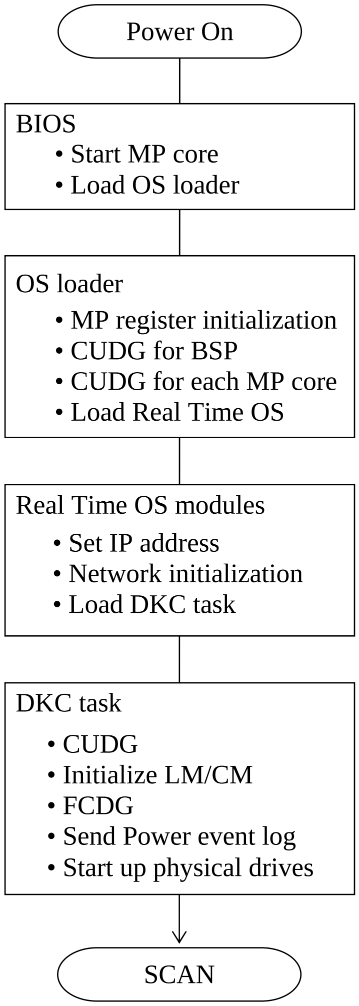


Fig. 4.1-1 IMPL Sequence

Copyright © 2012, Hitachi, Ltd.

4.2 (Blank)

Copyright © 2012, Hitachi, Ltd.

4.3 Planned Stop

When a power-off is specified by a maintenance personnel, this storage system checks for termination of tasks that are blocked or running on all logical devices. When all the tasks are terminated, this storage system disables the CHL and executes emergency destaging. If a track for which destaging fails (pinned track) occurs, this storage system stores the pin information in shared memory.
Subsequently, this storage system saves the configuration data and the pin information(which is used as hand-over information) in flash memory of MPBs, save all SM data (which is used as none-volatile power on) in SSD memory of CPCs. Then, sends Power Event Log to the SVP, notifies the hardware of the grant to turn off the power.

The hardware turns off main power when power-off grants for all processors are presented.

SVP	MP
	PS-off detected
	|
	Disable the channel Executes emergency destaging
	|
	Collect ORM information
	|
	Turn off drive power
	|
	Store SM data to SDD memory of CPCs
	|
	Store configuration data in FM of MPBs Store pin information in FM of MPBs
	|
←	Send Power Event Log to SVP
	|
	Grant PS-off
	DKC PS off

Copyright © 2012, Hitachi, Ltd.

5. Appendixes

5.1 Physical-Logical Device Matrixes

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (1/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-00	HDD00-00	00/00	000
	HDD00-01	00/01	001
	HDD00-02	00/02	002
	HDD00-03	00/03	003
	HDD00-04	00/04	004
	HDD00-05	00/05	005
	HDD00-06	00/06	006
	HDD00-07	00/07	007
	HDD00-08	00/08	008
	HDD00-09	00/09	009
	HDD00-10	00/0A	00A
	HDD00-11	00/0B	00B
	HDD00-12	00/0C	00C
	HDD00-13	00/0D	00D
	HDD00-14	00/0E	00E
	HDD00-15	00/0F	00F
	HDD00-16	00/10	010
	HDD00-17	00/11	011
	HDD00-18	00/12	012
	HDD00-19	00/13	013
	HDD00-20	00/14	014
	HDD00-21	00/15	015
	HDD00-22	00/16	016
	HDD00-23	00/17	017

 Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (2/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-01	HDD01-00	01/00	040
	HDD01-01	01/01	041
	HDD01-02	01/02	042
	HDD01-03	01/03	043
	HDD01-04	01/04	044
	HDD01-05	01/05	045
	HDD01-06	01/06	046
	HDD01-07	01/07	047
	HDD01-08	01/08	048
	HDD01-09	01/09	049
	HDD01-10	01/0A	04A
	HDD01-11	01/0B	04B
	HDD01-12	01/0C	04C
	HDD01-13	01/0D	04D
	HDD01-14	01/0E	04E
	HDD01-15	01/0F	04F
	HDD01-16	01/10	050
	HDD01-17	01/11	051
	HDD01-18	01/12	052
	HDD01-19	01/13	053
	HDD01-20	01/14	054
	HDD01-21	01/15	055
	HDD01-22	01/16	056
	HDD01-23	01/17	057

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (3/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-02	HDD02-00	02/00	080
	HDD02-01	02/01	081
	HDD02-02	02/02	082
	HDD02-03	02/03	083
	HDD02-04	02/04	084
	HDD02-05	02/05	085
	HDD02-06	02/06	086
	HDD02-07	02/07	087
	HDD02-08	02/08	088
	HDD02-09	02/09	089
	HDD02-10	02/0A	08A
	HDD02-11	02/0B	08B
	HDD02-12	02/0C	08C
	HDD02-13	02/0D	08D
	HDD02-14	02/0E	08E
	HDD02-15	02/0F	08F
	HDD02-16	02/10	090
	HDD02-17	02/11	091
	HDD02-18	02/12	092
	HDD02-19	02/13	093
	HDD02-20	02/14	094
	HDD02-21	02/15	095
	HDD02-22	02/16	096
	HDD02-23	02/17	097

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (4/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-03	HDD03-00	03/00	0C0
	HDD03-01	03/01	0C1
	HDD03-02	03/02	0C2
	HDD03-03	03/03	0C3
	HDD03-04	03/04	0C4
	HDD03-05	03/05	0C5
	HDD03-06	03/06	0C6
	HDD03-07	03/07	0C7
	HDD03-08	03/08	0C8
	HDD03-09	03/09	0C9
	HDD03-10	03/0A	0CA
	HDD03-11	03/0B	0CB
	HDD03-12	03/0C	0CC
	HDD03-13	03/0D	0CD
	HDD03-14	03/0E	0CE
	HDD03-15	03/0F	0CF
	HDD03-16	03/10	0D0
	HDD03-17	03/11	0D1
	HDD03-18	03/12	0D2
	HDD03-19	03/13	0D3
	HDD03-20	03/14	0D4
	HDD03-21	03/15	0D5
	HDD03-22	03/16	0D6
	HDD03-23	03/17	0D7

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (5/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-04	HDD04-00	04/00	100
	HDD04-01	04/01	101
	HDD04-02	04/02	102
	HDD04-03	04/03	103
	HDD04-04	04/04	104
	HDD04-05	04/05	105
	HDD04-06	04/06	106
	HDD04-07	04/07	107
	HDD04-08	04/08	108
	HDD04-09	04/09	109
	HDD04-10	04/0A	10A
	HDD04-11	04/0B	10B
	HDD04-12	04/0C	10C
	HDD04-13	04/0D	10D
	HDD04-14	04/0E	10E
	HDD04-15	04/0F	10F
	HDD04-16	04/10	110
	HDD04-17	04/11	111
	HDD04-18	04/12	112
	HDD04-19	04/13	113
	HDD04-20	04/14	114
	HDD04-21	04/15	115
	HDD04-22	04/16	116
	HDD04-23	04/17	117

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (6/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-05	HDD05-00	05/00	140
	HDD05-01	05/01	141
	HDD05-02	05/02	142
	HDD05-03	05/03	143
	HDD05-04	05/04	144
	HDD05-05	05/05	145
	HDD05-06	05/06	146
	HDD05-07	05/07	147
	HDD05-08	05/08	148
	HDD05-09	05/09	149
	HDD05-10	05/0A	14A
	HDD05-11	05/0B	14B
	HDD05-12	05/0C	14C
	HDD05-13	05/0D	14D
	HDD05-14	05/0E	14E
	HDD05-15	05/0F	14F
	HDD05-16	05/10	150
	HDD05-17	05/11	151
	HDD05-18	05/12	152
	HDD05-19	05/13	153
	HDD05-20	05/14	154
	HDD05-21	05/15	155
	HDD05-22	05/16	156
	HDD05-23	05/17	157

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (7/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-06	HDD06-00	06/00	180
	HDD06-01	06/01	181
	HDD06-02	06/02	182
	HDD06-03	06/03	183
	HDD06-04	06/04	184
	HDD06-05	06/05	185
	HDD06-06	06/06	186
	HDD06-07	06/07	187
	HDD06-08	06/08	188
	HDD06-09	06/09	189
	HDD06-10	06/0A	18A
	HDD06-11	06/0B	18B
	HDD06-12	06/0C	18C
	HDD06-13	06/0D	18D
	HDD06-14	06/0E	18E
	HDD06-15	06/0F	18F
	HDD06-16	06/10	190
	HDD06-17	06/11	191
	HDD06-18	06/12	192
	HDD06-19	06/13	193
	HDD06-20	06/14	194
	HDD06-21	06/15	195
	HDD06-22	06/16	196
	HDD06-23	06/17	197

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (8/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-07	HDD07-00	07/00	1C0
	HDD07-01	07/01	1C1
	HDD07-02	07/02	1C2
	HDD07-03	07/03	1C3
	HDD07-04	07/04	1C4
	HDD07-05	07/05	1C5
	HDD07-06	07/06	1C6
	HDD07-07	07/07	1C7
	HDD07-08	07/08	1C8
	HDD07-09	07/09	1C9
	HDD07-10	07/0A	1CA
	HDD07-11	07/0B	1CB
	HDD07-12	07/0C	1CC
	HDD07-13	07/0D	1CD
	HDD07-14	07/0E	1CE
	HDD07-15	07/0F	1CF
	HDD07-16	07/10	1D0
	HDD07-17	07/11	1D1
	HDD07-18	07/12	1D2
	HDD07-19	07/13	1D3
	HDD07-20	07/14	1D4
	HDD07-21	07/15	1D5
	HDD07-22	07/16	1D6
	HDD07-23	07/17	1D7

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (9/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-08	HDD08-00	08/00	200
	HDD08-01	08/01	201
	HDD08-02	08/02	202
	HDD08-03	08/03	203
	HDD08-04	08/04	204
	HDD08-05	08/05	205
	HDD08-06	08/06	206
	HDD08-07	08/07	207
	HDD08-08	08/08	208
	HDD08-09	08/09	209
	HDD08-10	08/0A	20A
	HDD08-11	08/0B	20B
	HDD08-12	08/0C	20C
	HDD08-13	08/0D	20D
	HDD08-14	08/0E	20E
	HDD08-15	08/0F	20F
	HDD08-16	08/10	210
	HDD08-17	08/11	211
	HDD08-18	08/12	212
	HDD08-19	08/13	213
	HDD08-20	08/14	214
	HDD08-21	08/15	215
	HDD08-22	08/16	216
	HDD08-23	08/17	217

 Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (10/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-09	HDD09-00	09/00	240
	HDD09-01	09/01	241
	HDD09-02	09/02	242
	HDD09-03	09/03	243
	HDD09-04	09/04	244
	HDD09-05	09/05	245
	HDD09-06	09/06	246
	HDD09-07	09/07	247
	HDD09-08	09/08	248
	HDD09-09	09/09	249
	HDD09-10	09/0A	24A
	HDD09-11	09/0B	24B
	HDD09-12	09/0C	24C
	HDD09-13	09/0D	24D
	HDD09-14	09/0E	24E
	HDD09-15	09/0F	24F
	HDD09-16	09/10	250
	HDD09-17	09/11	251
	HDD09-18	09/12	252
	HDD09-19	09/13	253
	HDD09-20	09/14	254
	HDD09-21	09/15	255
	HDD09-22	09/16	256
	HDD09-23	09/17	257

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (11/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-10	HDD10-00	0A/00	280
	HDD10-01	0A/01	281
	HDD10-02	0A/02	282
	HDD10-03	0A/03	283
	HDD10-04	0A/04	284
	HDD10-05	0A/05	285
	HDD10-06	0A/06	286
	HDD10-07	0A/07	287
	HDD10-08	0A/08	288
	HDD10-09	0A/09	289
	HDD10-10	0A/0A	28A
	HDD10-11	0A/0B	28B
	HDD10-12	0A/0C	28C
	HDD10-13	0A/0D	28D
	HDD10-14	0A/0E	28E
	HDD10-15	0A/0F	28F
	HDD10-16	0A/10	290
	HDD10-17	0A/11	291
	HDD10-18	0A/12	292
	HDD10-19	0A/13	293
	HDD10-20	0A/14	294
	HDD10-21	0A/15	295
	HDD10-22	0A/16	296
	HDD10-23	0A/17	297

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (12/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-11	HDD11-00	0B/00	2C0
	HDD11-01	0B/01	2C1
	HDD11-02	0B/02	2C2
	HDD11-03	0B/03	2C3
	HDD11-04	0B/04	2C4
	HDD11-05	0B/05	2C5
	HDD11-06	0B/06	2C6
	HDD11-07	0B/07	2C7
	HDD11-08	0B/08	2C8
	HDD11-09	0B/09	2C9
	HDD11-10	0B/0A	2CA
	HDD11-11	0B/0B	2CB
	HDD11-12	0B/0C	2CC
	HDD11-13	0B/0D	2CD
	HDD11-14	0B/0E	2CE
	HDD11-15	0B/0F	2CF
	HDD11-16	0B/10	2D0
	HDD11-17	0B/11	2D1
	HDD11-18	0B/12	2D2
	HDD11-19	0B/13	2D3
	HDD11-20	0B/14	2D4
	HDD11-21	0B/15	2D5
	HDD11-22	0B/16	2D6
	HDD11-23	0B/17	2D7

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (13/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-12	HDD12-00	0C/00	300
	HDD12-01	0C/01	301
	HDD12-02	0C/02	302
	HDD12-03	0C/03	303
	HDD12-04	0C/04	304
	HDD12-05	0C/05	305
	HDD12-06	0C/06	306
	HDD12-07	0C/07	307
	HDD12-08	0C/08	308
	HDD12-09	0C/09	309
	HDD12-10	0C/0A	30A
	HDD12-11	0C/0B	30B
	HDD12-12	0C/0C	30C
	HDD12-13	0C/0D	30D
	HDD12-14	0C/0E	30E
	HDD12-15	0C/0F	30F
	HDD12-16	0C/10	310
	HDD12-17	0C/11	311
	HDD12-18	0C/12	312
	HDD12-19	0C/13	313
	HDD12-20	0C/14	314
	HDD12-21	0C/15	315
	HDD12-22	0C/16	316
	HDD12-23	0C/17	317

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (14/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-13	HDD13-00	0D/00	340
	HDD13-01	0D/01	341
	HDD13-02	0D/02	342
	HDD13-03	0D/03	343
	HDD13-04	0D/04	344
	HDD13-05	0D/05	345
	HDD13-06	0D/06	346
	HDD13-07	0D/07	347
	HDD13-08	0D/08	348
	HDD13-09	0D/09	349
	HDD13-10	0D/0A	34A
	HDD13-11	0D/0B	34B
	HDD13-12	0D/0C	34C
	HDD13-13	0D/0D	34D
	HDD13-14	0D/0E	34E
	HDD13-15	0D/0F	34F
	HDD13-16	0D/10	350
	HDD13-17	0D/11	351
	HDD13-18	0D/12	352
	HDD13-19	0D/13	353
	HDD13-20	0D/14	354
	HDD13-21	0D/15	355
	HDD13-22	0D/16	356
	HDD13-23	0D/17	357

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (15/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-14	HDD14-00	0E/00	380
	HDD14-01	0E/01	381
	HDD14-02	0E/02	382
	HDD14-03	0E/03	383
	HDD14-04	0E/04	384
	HDD14-05	0E/05	385
	HDD14-06	0E/06	386
	HDD14-07	0E/07	387
	HDD14-08	0E/08	388
	HDD14-09	0E/09	389
	HDD14-10	0E/0A	38A
	HDD14-11	0E/0B	38B
	HDD14-12	0E/0C	38C
	HDD14-13	0E/0D	38D
	HDD14-14	0E/0E	38E
	HDD14-15	0E/0F	38F
	HDD14-16	0E/10	390
	HDD14-17	0E/11	391
	HDD14-18	0E/12	392
	HDD14-19	0E/13	393
	HDD14-20	0E/14	394
	HDD14-21	0E/15	395
	HDD14-22	0E/16	396
	HDD14-23	0E/17	397

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (16/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-15	HDD15-00	0F/00	3C0
	HDD15-01	0F/01	3C1
	HDD15-02	0F/02	3C2
	HDD15-03	0F/03	3C3
	HDD15-04	0F/04	3C4
	HDD15-05	0F/05	3C5
	HDD15-06	0F/06	3C6
	HDD15-07	0F/07	3C7
	HDD15-08	0F/08	3C8
	HDD15-09	0F/09	3C9
	HDD15-10	0F/0A	3CA
	HDD15-11	0F/0B	3CB
	HDD15-12	0F/0C	3CC
	HDD15-13	0F/0D	3CD
	HDD15-14	0F/0E	3CE
	HDD15-15	0F/0F	3CF
	HDD15-16	0F/10	3D0
	HDD15-17	0F/11	3D1
	HDD15-18	0F/12	3D2
	HDD15-19	0F/13	3D3
	HDD15-20	0F/14	3D4
	HDD15-21	0F/15	3D5
	HDD15-22	0F/16	3D6
	HDD15-23	0F/17	3D7

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (17/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-16	HDD16-00	10/00	400
	HDD16-01	10/01	401
	HDD16-02	10/02	402
	HDD16-03	10/03	403
	HDD16-04	10/04	404
	HDD16-05	10/05	405
	HDD16-06	10/06	406
	HDD16-07	10/07	407
	HDD16-08	10/08	408
	HDD16-09	10/09	409
	HDD16-10	10/0A	40A
	HDD16-11	10/0B	40B
	HDD16-12	10/0C	40C
	HDD16-13	10/0D	40D
	HDD16-14	10/0E	40E
	HDD16-15	10/0F	40F
	HDD16-16	10/10	410
	HDD16-17	10/11	411
	HDD16-18	10/12	412
	HDD16-19	10/13	413
	HDD16-20	10/14	414
	HDD16-21	10/15	415
	HDD16-22	10/16	416
	HDD16-23	10/17	417

 Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (18/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-17	HDD17-00	11/00	440
	HDD17-01	11/01	441
	HDD17-02	11/02	442
	HDD17-03	11/03	443
	HDD17-04	11/04	444
	HDD17-05	11/05	445
	HDD17-06	11/06	446
	HDD17-07	11/07	447
	HDD17-08	11/08	448
	HDD17-09	11/09	449
	HDD17-10	11/0A	44A
	HDD17-11	11/0B	44B
	HDD17-12	11/0C	44C
	HDD17-13	11/0D	44D
	HDD17-14	11/0E	44E
	HDD17-15	11/0F	44F
	HDD17-16	11/10	450
	HDD17-17	11/11	451
	HDD17-18	11/12	452
	HDD17-19	11/13	453
	HDD17-20	11/14	454
	HDD17-21	11/15	455
	HDD17-22	11/16	456
	HDD17-23	11/17	457

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (19/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-18	HDD18-00	12/00	480
	HDD18-01	12/01	481
	HDD18-02	12/02	482
	HDD18-03	12/03	483
	HDD18-04	12/04	484
	HDD18-05	12/05	485
	HDD18-06	12/06	486
	HDD18-07	12/07	487
	HDD18-08	12/08	488
	HDD18-09	12/09	489
	HDD18-10	12/0A	48A
	HDD18-11	12/0B	48B
	HDD18-12	12/0C	48C
	HDD18-13	12/0D	48D
	HDD18-14	12/0E	48E
	HDD18-15	12/0F	48F
	HDD18-16	12/10	490
	HDD18-17	12/11	491
	HDD18-18	12/12	492
	HDD18-19	12/13	493
	HDD18-20	12/14	494
	HDD18-21	12/15	495
	HDD18-22	12/16	496
	HDD18-23	12/17	497

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (20/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-19	HDD19-00	13/00	4C0
	HDD19-01	13/01	4C1
	HDD19-02	13/02	4C2
	HDD19-03	13/03	4C3
	HDD19-04	13/04	4C4
	HDD19-05	13/05	4C5
	HDD19-06	13/06	4C6
	HDD19-07	13/07	4C7
	HDD19-08	13/08	4C8
	HDD19-09	13/09	4C9
	HDD19-10	13/0A	4CA
	HDD19-11	13/0B	4CB
	HDD19-12	13/0C	4CC
	HDD19-13	13/0D	4CD
	HDD19-14	13/0E	4CE
	HDD19-15	13/0F	4CF
	HDD19-16	13/10	4D0
	HDD19-17	13/11	4D1
	HDD19-18	13/12	4D2
	HDD19-19	13/13	4D3
	HDD19-20	13/14	4D4
	HDD19-21	13/15	4D5
	HDD19-22	13/16	4D6
	HDD19-23	13/17	4D7

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (21/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-20	HDD20-00	14/00	500
	HDD20-01	14/01	501
	HDD20-02	14/02	502
	HDD20-03	14/03	503
	HDD20-04	14/04	504
	HDD20-05	14/05	505
	HDD20-06	14/06	506
	HDD20-07	14/07	507
	HDD20-08	14/08	508
	HDD20-09	14/09	509
	HDD20-10	14/0A	50A
	HDD20-11	14/0B	50B
	HDD20-12	14/0C	50C
	HDD20-13	14/0D	50D
	HDD20-14	14/0E	50E
	HDD20-15	14/0F	50F
	HDD20-16	14/10	510
	HDD20-17	14/11	511
	HDD20-18	14/12	512
	HDD20-19	14/13	513
	HDD20-20	14/14	514
	HDD20-21	14/15	515
	HDD20-22	14/16	516
	HDD20-23	14/17	517

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (22/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-21	HDD21-00	15/00	540
	HDD21-01	15/01	541
	HDD21-02	15/02	542
	HDD21-03	15/03	543
	HDD21-04	15/04	544
	HDD21-05	15/05	545
	HDD21-06	15/06	546
	HDD21-07	15/07	547
	HDD21-08	15/08	548
	HDD21-09	15/09	549
	HDD21-10	15/0A	54A
	HDD21-11	15/0B	54B
	HDD21-12	15/0C	54C
	HDD21-13	15/0D	54D
	HDD21-14	15/0E	54E
	HDD21-15	15/0F	54F
	HDD21-16	15/10	550
	HDD21-17	15/11	551
	HDD21-18	15/12	552
	HDD21-19	15/13	553
	HDD21-20	15/14	554
	HDD21-21	15/15	555
	HDD21-22	15/16	556
	HDD21-23	15/17	557

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (23/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-22	HDD22-00	16/00	580
	HDD22-01	16/01	581
	HDD22-02	16/02	582
	HDD22-03	16/03	583
	HDD22-04	16/04	584
	HDD22-05	16/05	585
	HDD22-06	16/06	586
	HDD22-07	16/07	587
	HDD22-08	16/08	588
	HDD22-09	16/09	589
	HDD22-10	16/0A	58A
	HDD22-11	16/0B	58B
	HDD22-12	16/0C	58C
	HDD22-13	16/0D	58D
	HDD22-14	16/0E	58E
	HDD22-15	16/0F	58F
	HDD22-16	16/10	590
	HDD22-17	16/11	591
	HDD22-18	16/12	592
	HDD22-19	16/13	593
	HDD22-20	16/14	594
	HDD22-21	16/15	595
	HDD22-22	16/16	596
	HDD22-23	16/17	597

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (24/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-23	HDD23-00	17/00	5C0
	HDD23-01	17/01	5C1
	HDD23-02	17/02	5C2
	HDD23-03	17/03	5C3
	HDD23-04	17/04	5C4
	HDD23-05	17/05	5C5
	HDD23-06	17/06	5C6
	HDD23-07	17/07	5C7
	HDD23-08	17/08	5C8
	HDD23-09	17/09	5C9
	HDD23-10	17/0A	5CA
	HDD23-11	17/0B	5CB
	HDD23-12	17/0C	5CC
	HDD23-13	17/0D	5CD
	HDD23-14	17/0E	5CE
	HDD23-15	17/0F	5CF
	HDD23-16	17/10	5D0
	HDD23-17	17/11	5D1
	HDD23-18	17/12	5D2
	HDD23-19	17/13	5D3
	HDD23-20	17/14	5D4
	HDD23-21	17/15	5D5
	HDD23-22	17/16	5D6
	HDD23-23	17/17	5D7

 Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (25/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-24	HDD24-00	18/00	600
	HDD24-01	18/01	601
	HDD24-02	18/02	602
	HDD24-03	18/03	603
	HDD24-04	18/04	604
	HDD24-05	18/05	605
	HDD24-06	18/06	606
	HDD24-07	18/07	607
	HDD24-08	18/08	608
	HDD24-09	18/09	609
	HDD24-10	18/0A	60A
	HDD24-11	18/0B	60B
	HDD24-12	18/0C	60C
	HDD24-13	18/0D	60D
	HDD24-14	18/0E	60E
	HDD24-15	18/0F	60F
	HDD24-16	18/10	610
	HDD24-17	18/11	611
	HDD24-18	18/12	612
	HDD24-19	18/13	613
	HDD24-20	18/14	614
	HDD24-21	18/15	615
	HDD24-22	18/16	616
	HDD24-23	18/17	617

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (26/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-25	HDD25-00	19/00	640
	HDD25-01	19/01	641
	HDD25-02	19/02	642
	HDD25-03	19/03	643
	HDD25-04	19/04	644
	HDD25-05	19/05	645
	HDD25-06	19/06	646
	HDD25-07	19/07	647
	HDD25-08	19/08	648
	HDD25-09	19/09	649
	HDD25-10	19/0A	64A
	HDD25-11	19/0B	64B
	HDD25-12	19/0C	64C
	HDD25-13	19/0D	64D
	HDD25-14	19/0E	64E
	HDD25-15	19/0F	64F
	HDD25-16	19/10	650
	HDD25-17	19/11	651
	HDD25-18	19/12	652
	HDD25-19	19/13	653
	HDD25-20	19/14	654
	HDD25-21	19/15	655
	HDD25-22	19/16	656
	HDD25-23	19/17	657

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (27/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-26	HDD26-00	1A/00	680
	HDD26-01	1A/01	681
	HDD26-02	1A/02	682
	HDD26-03	1A/03	683
	HDD26-04	1A/04	684
	HDD26-05	1A/05	685
	HDD26-06	1A/06	686
	HDD26-07	1A/07	687
	HDD26-08	1A/08	688
	HDD26-09	1A/09	689
	HDD26-10	1A/0A	68A
	HDD26-11	1A/0B	68B
	HDD26-12	1A/0C	68C
	HDD26-13	1A/0D	68D
	HDD26-14	1A/0E	68E
	HDD26-15	1A/0F	68F
	HDD26-16	1A/10	690
	HDD26-17	1A/11	691
	HDD26-18	1A/12	692
	HDD26-19	1A/13	693
	HDD26-20	1A/14	694
	HDD26-21	1A/15	695
	HDD26-22	1A/16	696
	HDD26-23	1A/17	697

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (28/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-27	HDD27-00	1B/00	6C0
	HDD27-01	1B/01	6C1
	HDD27-02	1B/02	6C2
	HDD27-03	1B/03	6C3
	HDD27-04	1B/04	6C4
	HDD27-05	1B/05	6C5
	HDD27-06	1B/06	6C6
	HDD27-07	1B/07	6C7
	HDD27-08	1B/08	6C8
	HDD27-09	1B/09	6C9
	HDD27-10	1B/0A	6CA
	HDD27-11	1B/0B	6CB
	HDD27-12	1B/0C	6CC
	HDD27-13	1B/0D	6CD
	HDD27-14	1B/0E	6CE
	HDD27-15	1B/0F	6CF
	HDD27-16	1B/10	6D0
	HDD27-17	1B/11	6D1
	HDD27-18	1B/12	6D2
	HDD27-19	1B/13	6D3
	HDD27-20	1B/14	6D4
	HDD27-21	1B/15	6D5
	HDD27-22	1B/16	6D6
	HDD27-23	1B/17	6D7

 Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (29/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-28	HDD28-00	1C/00	700
	HDD28-01	1C/01	701
	HDD28-02	1C/02	702
	HDD28-03	1C/03	703
	HDD28-04	1C/04	704
	HDD28-05	1C/05	705
	HDD28-06	1C/06	706
	HDD28-07	1C/07	707
	HDD28-08	1C/08	708
	HDD28-09	1C/09	709
	HDD28-10	1C/0A	70A
	HDD28-11	1C/0B	70B
	HDD28-12	1C/0C	70C
	HDD28-13	1C/0D	70D
	HDD28-14	1C/0E	70E
	HDD28-15	1C/0F	70F
	HDD28-16	1C/10	710
	HDD28-17	1C/11	711
	HDD28-18	1C/12	712
	HDD28-19	1C/13	713
	HDD28-20	1C/14	714
	HDD28-21	1C/15	715
	HDD28-22	1C/16	716
	HDD28-23	1C/17	717

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (30/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-29	HDD29-00	1D/00	740
	HDD29-01	1D/01	741
	HDD29-02	1D/02	742
	HDD29-03	1D/03	743
	HDD29-04	1D/04	744
	HDD29-05	1D/05	745
	HDD29-06	1D/06	746
	HDD29-07	1D/07	747
	HDD29-08	1D/08	748
	HDD29-09	1D/09	749
	HDD29-10	1D/0A	74A
	HDD29-11	1D/0B	74B
	HDD29-12	1D/0C	74C
	HDD29-13	1D/0D	74D
	HDD29-14	1D/0E	74E
	HDD29-15	1D/0F	74F
	HDD29-16	1D/10	750
	HDD29-17	1D/11	751
	HDD29-18	1D/12	752
	HDD29-19	1D/13	753
	HDD29-20	1D/14	754
	HDD29-21	1D/15	755
	HDD29-22	1D/16	756
	HDD29-23	1D/17	757

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (31/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-30	HDD30-00	1E/00	780
	HDD30-01	1E/01	781
	HDD30-02	1E/02	782
	HDD30-03	1E/03	783
	HDD30-04	1E/04	784
	HDD30-05	1E/05	785
	HDD30-06	1E/06	786
	HDD30-07	1E/07	787
	HDD30-08	1E/08	788
	HDD30-09	1E/09	789
	HDD30-10	1E/0A	78A
	HDD30-11	1E/0B	78B
	HDD30-12	1E/0C	78C
	HDD30-13	1E/0D	78D
	HDD30-14	1E/0E	78E
	HDD30-15	1E/0F	78F
	HDD30-16	1E/10	790
	HDD30-17	1E/11	791
	HDD30-18	1E/12	792
	HDD30-19	1E/13	793
	HDD30-20	1E/14	794
	HDD30-21	1E/15	795
	HDD30-22	1E/16	796
	HDD30-23	1E/17	797

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (32/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-31	HDD31-00	1F/00	7C0
	HDD31-01	1F/01	7C1
	HDD31-02	1F/02	7C2
	HDD31-03	1F/03	7C3
	HDD31-04	1F/04	7C4
	HDD31-05	1F/05	7C5
	HDD31-06	1F/06	7C6
	HDD31-07	1F/07	7C7
	HDD31-08	1F/08	7C8
	HDD31-09	1F/09	7C9
	HDD31-10	1F/0A	7CA
	HDD31-11	1F/0B	7CB
	HDD31-12	1F/0C	7CC
	HDD31-13	1F/0D	7CD
	HDD31-14	1F/0E	7CE
	HDD31-15	1F/0F	7CF
	HDD31-16	1F/10	7D0
	HDD31-17	1F/11	7D1
	HDD31-18	1F/12	7D2
	HDD31-19	1F/13	7D3
	HDD31-20	1F/14	7D4
	HDD31-21	1F/15	7D5
	HDD31-22	1F/16	7D6
	HDD31-23	1F/17	7D7

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (33/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-32	HDD32-00	20/00	800
	HDD32-01	20/01	801
	HDD32-02	20/02	802
	HDD32-03	20/03	803
	HDD32-04	20/04	804
	HDD32-05	20/05	805
	HDD32-06	20/06	806
	HDD32-07	20/07	807
	HDD32-08	20/08	808
	HDD32-09	20/09	809
	HDD32-10	20/0A	80A
	HDD32-11	20/0B	80B
	HDD32-12	20/0C	80C
	HDD32-13	20/0D	80D
	HDD32-14	20/0E	80E
	HDD32-15	20/0F	80F
	HDD32-16	20/10	810
	HDD32-17	20/11	811
	HDD32-18	20/12	812
	HDD32-19	20/13	813
	HDD32-20	20/14	814
	HDD32-21	20/15	815
	HDD32-22	20/16	816
	HDD32-23	20/17	817

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (34/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-33	HDD33-00	21/00	840
	HDD33-01	21/01	841
	HDD33-02	21/02	842
	HDD33-03	21/03	843
	HDD33-04	21/04	844
	HDD33-05	21/05	845
	HDD33-06	21/06	846
	HDD33-07	21/07	847
	HDD33-08	21/08	848
	HDD33-09	21/09	849
	HDD33-10	21/0A	84A
	HDD33-11	21/0B	84B
	HDD33-12	21/0C	84C
	HDD33-13	21/0D	84D
	HDD33-14	21/0E	84E
	HDD33-15	21/0F	84F
	HDD33-16	21/10	850
	HDD33-17	21/11	851
	HDD33-18	21/12	852
	HDD33-19	21/13	853
	HDD33-20	21/14	854
	HDD33-21	21/15	855
	HDD33-22	21/16	856
	HDD33-23	21/17	857

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (35/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-34	HDD34-00	22/00	880
	HDD34-01	22/01	881
	HDD34-02	22/02	882
	HDD34-03	22/03	883
	HDD34-04	22/04	884
	HDD34-05	22/05	885
	HDD34-06	22/06	886
	HDD34-07	22/07	887
	HDD34-08	22/08	888
	HDD34-09	22/09	889
	HDD34-10	22/0A	88A
	HDD34-11	22/0B	88B
	HDD34-12	22/0C	88C
	HDD34-13	22/0D	88D
	HDD34-14	22/0E	88E
	HDD34-15	22/0F	88F
	HDD34-16	22/10	890
	HDD34-17	22/11	891
	HDD34-18	22/12	892
	HDD34-19	22/13	893
	HDD34-20	22/14	894
	HDD34-21	22/15	895
	HDD34-22	22/16	896
	HDD34-23	22/17	897

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (36/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-35	HDD35-00	23/00	8C0
	HDD35-01	23/01	8C1
	HDD35-02	23/02	8C2
	HDD35-03	23/03	8C3
	HDD35-04	23/04	8C4
	HDD35-05	23/05	8C5
	HDD35-06	23/06	8C6
	HDD35-07	23/07	8C7
	HDD35-08	23/08	8C8
	HDD35-09	23/09	8C9
	HDD35-10	23/0A	8CA
	HDD35-11	23/0B	8CB
	HDD35-12	23/0C	8CC
	HDD35-13	23/0D	8CD
	HDD35-14	23/0E	8CE
	HDD35-15	23/0F	8CF
	HDD35-16	23/10	8D0
	HDD35-17	23/11	8D1
	HDD35-18	23/12	8D2
	HDD35-19	23/13	8D3
	HDD35-20	23/14	8D4
	HDD35-21	23/15	8D5
	HDD35-22	23/16	8D6
	HDD35-23	23/17	8D7

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (37/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-36	HDD36-00	24/00	900
	HDD36-01	24/01	901
	HDD36-02	24/02	902
	HDD36-03	24/03	903
	HDD36-04	24/04	904
	HDD36-05	24/05	905
	HDD36-06	24/06	906
	HDD36-07	24/07	907
	HDD36-08	24/08	908
	HDD36-09	24/09	909
	HDD36-10	24/0A	90A
	HDD36-11	24/0B	90B
	HDD36-12	24/0C	90C
	HDD36-13	24/0D	90D
	HDD36-14	24/0E	90E
	HDD36-15	24/0F	90F
	HDD36-16	24/10	910
	HDD36-17	24/11	911
	HDD36-18	24/12	912
	HDD36-19	24/13	913
	HDD36-20	24/14	914
	HDD36-21	24/15	915
	HDD36-22	24/16	916
	HDD36-23	24/17	917

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (38/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-37	HDD37-00	25/00	940
	HDD37-01	25/01	941
	HDD37-02	25/02	942
	HDD37-03	25/03	943
	HDD37-04	25/04	944
	HDD37-05	25/05	945
	HDD37-06	25/06	946
	HDD37-07	25/07	947
	HDD37-08	25/08	948
	HDD37-09	25/09	949
	HDD37-10	25/0A	94A
	HDD37-11	25/0B	94B
	HDD37-12	25/0C	94C
	HDD37-13	25/0D	94D
	HDD37-14	25/0E	94E
	HDD37-15	25/0F	94F
	HDD37-16	25/10	950
	HDD37-17	25/11	951
	HDD37-18	25/12	952
	HDD37-19	25/13	953
	HDD37-20	25/14	954
	HDD37-21	25/15	955
	HDD37-22	25/16	956
	HDD37-23	25/17	957

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (39/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-38	HDD38-00	26/00	980
	HDD38-01	26/01	981
	HDD38-02	26/02	982
	HDD38-03	26/03	983
	HDD38-04	26/04	984
	HDD38-05	26/05	985
	HDD38-06	26/06	986
	HDD38-07	26/07	987
	HDD38-08	26/08	988
	HDD38-09	26/09	989
	HDD38-10	26/0A	98A
	HDD38-11	26/0B	98B
	HDD38-12	26/0C	98C
	HDD38-13	26/0D	98D
	HDD38-14	26/0E	98E
	HDD38-15	26/0F	98F
	HDD38-16	26/10	990
	HDD38-17	26/11	991
	HDD38-18	26/12	992
	HDD38-19	26/13	993
	HDD38-20	26/14	994
	HDD38-21	26/15	995
	HDD38-22	26/16	996
	HDD38-23	26/17	997

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (40/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-39	HDD39-00	27/00	9C0
	HDD39-01	27/01	9C1
	HDD39-02	27/02	9C2
	HDD39-03	27/03	9C3
	HDD39-04	27/04	9C4
	HDD39-05	27/05	9C5
	HDD39-06	27/06	9C6
	HDD39-07	27/07	9C7
	HDD39-08	27/08	9C8
	HDD39-09	27/09	9C9
	HDD39-10	27/0A	9CA
	HDD39-11	27/0B	9CB
	HDD39-12	27/0C	9CC
	HDD39-13	27/0D	9CD
	HDD39-14	27/0E	9CE
	HDD39-15	27/0F	9CF
	HDD39-16	27/10	9D0
	HDD39-17	27/11	9D1
	HDD39-18	27/12	9D2
	HDD39-19	27/13	9D3
	HDD39-20	27/14	9D4
	HDD39-21	27/15	9D5
	HDD39-22	27/16	9D6
	HDD39-23	27/17	9D7

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (41/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-40	HDD40-00	28/00	A00
	HDD40-01	28/01	A01
	HDD40-02	28/02	A02
	HDD40-03	28/03	A03
	HDD40-04	28/04	A04
	HDD40-05	28/05	A05
	HDD40-06	28/06	A06
	HDD40-07	28/07	A07
	HDD40-08	28/08	A08
	HDD40-09	28/09	A09
	HDD40-10	28/0A	A0A
	HDD40-11	28/0B	A0B
	HDD40-12	28/0C	A0C
	HDD40-13	28/0D	A0D
	HDD40-14	28/0E	A0E
	HDD40-15	28/0F	A0F
	HDD40-16	28/10	A10
	HDD40-17	28/11	A11
	HDD40-18	28/12	A12
	HDD40-19	28/13	A13
	HDD40-20	28/14	A14
	HDD40-21	28/15	A15
	HDD40-22	28/16	A16
	HDD40-23	28/17	A17

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (42/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-41	HDD41-00	29/00	A40
	HDD41-01	29/01	A41
	HDD41-02	29/02	A42
	HDD41-03	29/03	A43
	HDD41-04	29/04	A44
	HDD41-05	29/05	A45
	HDD41-06	29/06	A46
	HDD41-07	29/07	A47
	HDD41-08	29/08	A48
	HDD41-09	29/09	A49
	HDD41-10	29/0A	A4A
	HDD41-11	29/0B	A4B
	HDD41-12	29/0C	A4C
	HDD41-13	29/0D	A4D
	HDD41-14	29/0E	A4E
	HDD41-15	29/0F	A4F
	HDD41-16	29/10	A50
	HDD41-17	29/11	A51
	HDD41-18	29/12	A52
	HDD41-19	29/13	A53
	HDD41-20	29/14	A54
	HDD41-21	29/15	A55
	HDD41-22	29/16	A56
	HDD41-23	29/17	A57

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (43/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-42	HDD42-00	2A/00	A80
	HDD42-01	2A/01	A81
	HDD42-02	2A/02	A82
	HDD42-03	2A/03	A83
	HDD42-04	2A/04	A84
	HDD42-05	2A/05	A85
	HDD42-06	2A/06	A86
	HDD42-07	2A/07	A87
	HDD42-08	2A/08	A88
	HDD42-09	2A/09	A89
	HDD42-10	2A/0A	A8A
	HDD42-11	2A/0B	A8B
	HDD42-12	2A/0C	A8C
	HDD42-13	2A/0D	A8D
	HDD42-14	2A/0E	A8E
	HDD42-15	2A/0F	A8F
	HDD42-16	2A/10	A90
	HDD42-17	2A/11	A91
	HDD42-18	2A/12	A92
	HDD42-19	2A/13	A93
	HDD42-20	2A/14	A94
	HDD42-21	2A/15	A95
	HDD42-22	2A/16	A96
	HDD42-23	2A/17	A97

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (44/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-43	HDD43-00	2B/00	AC0
	HDD43-01	2B/01	AC1
	HDD43-02	2B/02	AC2
	HDD43-03	2B/03	AC3
	HDD43-04	2B/04	AC4
	HDD43-05	2B/05	AC5
	HDD43-06	2B/06	AC6
	HDD43-07	2B/07	AC7
	HDD43-08	2B/08	AC8
	HDD43-09	2B/09	AC9
	HDD43-10	2B/0A	ACA
	HDD43-11	2B/0B	ACB
	HDD43-12	2B/0C	ACC
	HDD43-13	2B/0D	ACD
	HDD43-14	2B/0E	ACE
	HDD43-15	2B/0F	ACF
	HDD43-16	2B/10	AD0
	HDD43-17	2B/11	AD1
	HDD43-18	2B/12	AD2
	HDD43-19	2B/13	AD3
	HDD43-20	2B/14	AD4
	HDD43-21	2B/15	AD5
	HDD43-22	2B/16	AD6
	HDD43-23	2B/17	AD7

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (45/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-44	HDD44-00	2C/00	B00
	HDD44-01	2C/01	B01
	HDD44-02	2C/02	B02
	HDD44-03	2C/03	B03
	HDD44-04	2C/04	B04
	HDD44-05	2C/05	B05
	HDD44-06	2C/06	B06
	HDD44-07	2C/07	B07
	HDD44-08	2C/08	B08
	HDD44-09	2C/09	B09
	HDD44-10	2C/0A	B0A
	HDD44-11	2C/0B	B0B
	HDD44-12	2C/0C	B0C
	HDD44-13	2C/0D	B0D
	HDD44-14	2C/0E	B0E
	HDD44-15	2C/0F	B0F
	HDD44-16	2C/10	B10
	HDD44-17	2C/11	B11
	HDD44-18	2C/12	B12
	HDD44-19	2C/13	B13
	HDD44-20	2C/14	B14
	HDD44-21	2C/15	B15
	HDD44-22	2C/16	B16
	HDD44-23	2C/17	B17

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (46/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-45	HDD45-00	2D/00	B40
	HDD45-01	2D/01	B41
	HDD45-02	2D/02	B42
	HDD45-03	2D/03	B43
	HDD45-04	2D/04	B44
	HDD45-05	2D/05	B45
	HDD45-06	2D/06	B46
	HDD45-07	2D/07	B47
	HDD45-08	2D/08	B48
	HDD45-09	2D/09	B49
	HDD45-10	2D/0A	B4A
	HDD45-11	2D/0B	B4B
	HDD45-12	2D/0C	B4C
	HDD45-13	2D/0D	B4D
	HDD45-14	2D/0E	B4E
	HDD45-15	2D/0F	B4F
	HDD45-16	2D/10	B50
	HDD45-17	2D/11	B51
	HDD45-18	2D/12	B52
	HDD45-19	2D/13	B53
	HDD45-20	2D/14	B54
	HDD45-21	2D/15	B55
	HDD45-22	2D/16	B56
	HDD45-23	2D/17	B57

Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (47/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-46	HDD46-00	2E/00	B80
	HDD46-01	2E/01	B81
	HDD46-02	2E/02	B82
	HDD46-03	2E/03	B83
	HDD46-04	2E/04	B84
	HDD46-05	2E/05	B85
	HDD46-06	2E/06	B86
	HDD46-07	2E/07	B87
	HDD46-08	2E/08	B88
	HDD46-09	2E/09	B89
	HDD46-10	2E/0A	B8A
	HDD46-11	2E/0B	B8B
	HDD46-12	2E/0C	B8C
	HDD46-13	2E/0D	B8D
	HDD46-14	2E/0E	B8E
	HDD46-15	2E/0F	B8F
	HDD46-16	2E/10	B90
	HDD46-17	2E/11	B91
	HDD46-18	2E/12	B92
	HDD46-19	2E/13	B93
	HDD46-20	2E/14	B94
	HDD46-21	2E/15	B95
	HDD46-22	2E/16	B96
	HDD46-23	2E/17	B97

 Copyright © 2012, Hitachi, Ltd.

RELATIONSHIP BETWEEN DISK DRIVE# AND PARITY GROUP# (48/48)

Drive Box Number	Disk Drive Number	C#/R#	Reference Code C#/R#(3-digit Display)
DB-47	HDD47-00	2F/00	BC0
	HDD47-01	2F/01	BC1
	HDD47-02	2F/02	BC2
	HDD47-03	2F/03	BC3
	HDD47-04	2F/04	BC4
	HDD47-05	2F/05	BC5
	HDD47-06	2F/06	BC6
	HDD47-07	2F/07	BC7
	HDD47-08	2F/08	BC8
	HDD47-09	2F/09	BC9
	HDD47-10	2F/0A	BCA
	HDD47-11	2F/0B	BCB
	HDD47-12	2F/0C	BCC
	HDD47-13	2F/0D	BCD
	HDD47-14	2F/0E	BCE
	HDD47-15	2F/0F	BCF
	HDD47-16	2F/10	BD0
	HDD47-17	2F/11	BD1
	HDD47-18	2F/12	BD2
	HDD47-19	2F/13	BD3
	HDD47-20	2F/14	BD4
	HDD47-21	2F/15	BD5
	HDD47-22	2F/16	BD6
	HDD47-23	2F/17	BD7

5.2 Comparison of pair status on SVP, Web Console, RAID Manager

Table.5.2-1 Comparison of pair status on SVP, Web Console, RAID Manager

No.	Event	Status on RAID Manager	Status on SVP, Web Console
1	Simplex Volume	P-VOL: SMPL S-VOL: SMPL	P-VOL: SMPL S-VOL: SMPL
2	Copying LUSE Volume Partly completed (SYNC only)	P-VOL: PDUB S-VOL: PDUB	P-VOL: PDUB S-VOL: PDUB
3	Copying Volume	P-VOL: COPY S-VOL: COPY	P-VOL: COPY S-VOL: COPY
4	Pair volume	P-VOL: PAIR S-VOL: PAIR	P-VOL: PAIR S-VOL: PAIR
5	Pairsplit operation to P-VOL	P-VOL: PSUS S-VOL: SSUS	P-VOL: PSUS (S-VOL by operator) S-VOL: PSUS (S-VOL by operator)/ SSUS
6	Pairsplit operation to S-VOL	P-VOL: PSUS S-VOL: PSUS	P-VOL: PSUS (S-VOL by operator) S-VOL: PSUS (S-VOL by operator)
7	Pairsplit -P operation (*1) (P-VOL failure, SYNC only)	P-VOL: PSUS S-VOL: SSUS	P-VOL: PSUS (P-VOL by operator) S-VOL: PSUS (by MCU)/SSUS
8	Pairsplit -R operation (*1)	P-VOL: PSUS S-VOL: SMPL	P-VOL: PSUS(Delete pair to RCU) S-VOL: SMPL
9	P-VOL Suspend (failure)	P-VOL: PSUE S-VOL: SSUS	P-VOL: PSUE (S-VOL failure) S-VOL: PSUE (S-VOL failure)/ SSUS
10	S-VOL Suspend (failure)	P-VOL: PSUE S-VOL: PSUE	P-VOL: PSUE (S-VOL failure) S-VOL: PSUE (S-VOL failure)
11	PS ON failure	P-VOL: PSUE S-VOL: —	P-VOL: PSUE (MCU IMPL) S-VOL: —
12	Copy failure (P-VOL failure)	P-VOL: PSUE S-VOL: SSUS	P-VOL: PSUE (Initial copy failed) S-VOL: PSUE (Initial copy failed)/ SSUS
13	Copy failure (S-VOL failure)	P-VOL: PSUE S-VOL: PSUE	P-VOL: PSUE (Initial copy failed) S-VOL: PSUE (Initial copy failed)
14	RCU accepted the notification of MCU's P/S-OFF	P-VOL: — S-VOL: SSUS	P-VOL: — S-VOL: PSUE (MCU P/S OFF)/ SSUS
15	MCU detected the obstacle of RCU	P-VOL: PSUE S-VOL: PSUE	P-VOL: PSUS (by RCU)/PSUE S-VOL: PSUE (S-VOL failure)

*1: Operation on RAID Manager

Copyright © 2012, Hitachi, Ltd.

5.3 Parts number of MAIN Blade, CHB/DKB, MPB

Table 5.3-1 Relationship between MAIN Blade and DMA#, DRR#

Location	MAIN Blade#	DMA#	DRR#
MAIN1	x00	x00 - x05	x00 - x03
MAIN2	x01	x08 - x0D	x04 - x07

Table 5.3-2 Relationship between CHB/DKB and Port#, SASCTL#

Location	Channel Port#	SASCTL#	SASPort#
CHB-1A	x00 - x03	-	-
CHB-1B	x04 - x07	-	-
CHB-1C	x08 - x0B	-	-
CHB-1D	x0C - x0F	-	-
CHB-1E	x10 - x13	-	-
CHB-1F	x14 - x17	-	-
DKB-1E	-	x00	x00 - x01
DKB-1F	-	x01	x02 - x03
CHB-2A	x20 - x23	-	-
CHB-2B	x24 - x27	-	-
CHB-2C	x28 - x2B	-	-
CHB-2D	x2C - x2F	-	-
CHB-2E	x30 – x33	-	-
CHB-2F	x34 – x37	-	-
DKB-2E	-	x02	x04 - x05
DKB-2F	-	x03	x06 - x07

Table 5.3-3 Relationship between MP Blade and MPB#, MP#

Location	MPB#	MPU#	MP# (MP core#)
MPB1	x00	x00 - x01	x00 - x07
MPB2	x01	x02 - x03	x08 - x0F

Copyright © 2012, Hitachi, Ltd.

5.4 Connection Diagram of DKC