US 20040193760A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2004/0193760 A1

Matsunami et al. (43) Pub. Date: Sep. 30, 2004

(54) STORAGE DEVICE

(75) Inventors: **Naoto Matsunami**, Hayama (JP); **Koji Sonoda**, Sagamihara (JP); **Akira Yamamoto**, Sagamihara (JP); **Masafumi Nozawa**, Odawara (JP); **Masaaki Iwasaki**, Tachikawa (JP)

Correspondence Address:
HOGAN & HARTSON L.L.P.
500 S. GRAND AVENUE
SUITE 1900
LOS ANGELES, CA 90071-2611 (US)

(73) Assignee: **HITACHI, LTD.**

(21) Appl. No.: **10/775,886**

(22) Filed: **Feb. 10, 2004**

(30) **Foreign Application Priority Data**

Mar. 27, 2003 (JP) 2003-086828

Publication Classification

(51) **Int. Cl.**[7] **G06F 3/00**

(52) **U.S. Cl.** **710/36**

(57) **ABSTRACT**

A storage device is provided with a file I/O interface control device and a plurality of disk pools. The file I/O interface control device sets one of a plurality of storage hierarchies defining storage classes, respectively, for each of LUs within the disk pools, thereby forming a file system in each of the LUs. The file I/O interface control device migrates at least one of the files from one of the LUs to another one of the LUs of an optimal storage class, based on static properties and dynamic properties of each file.

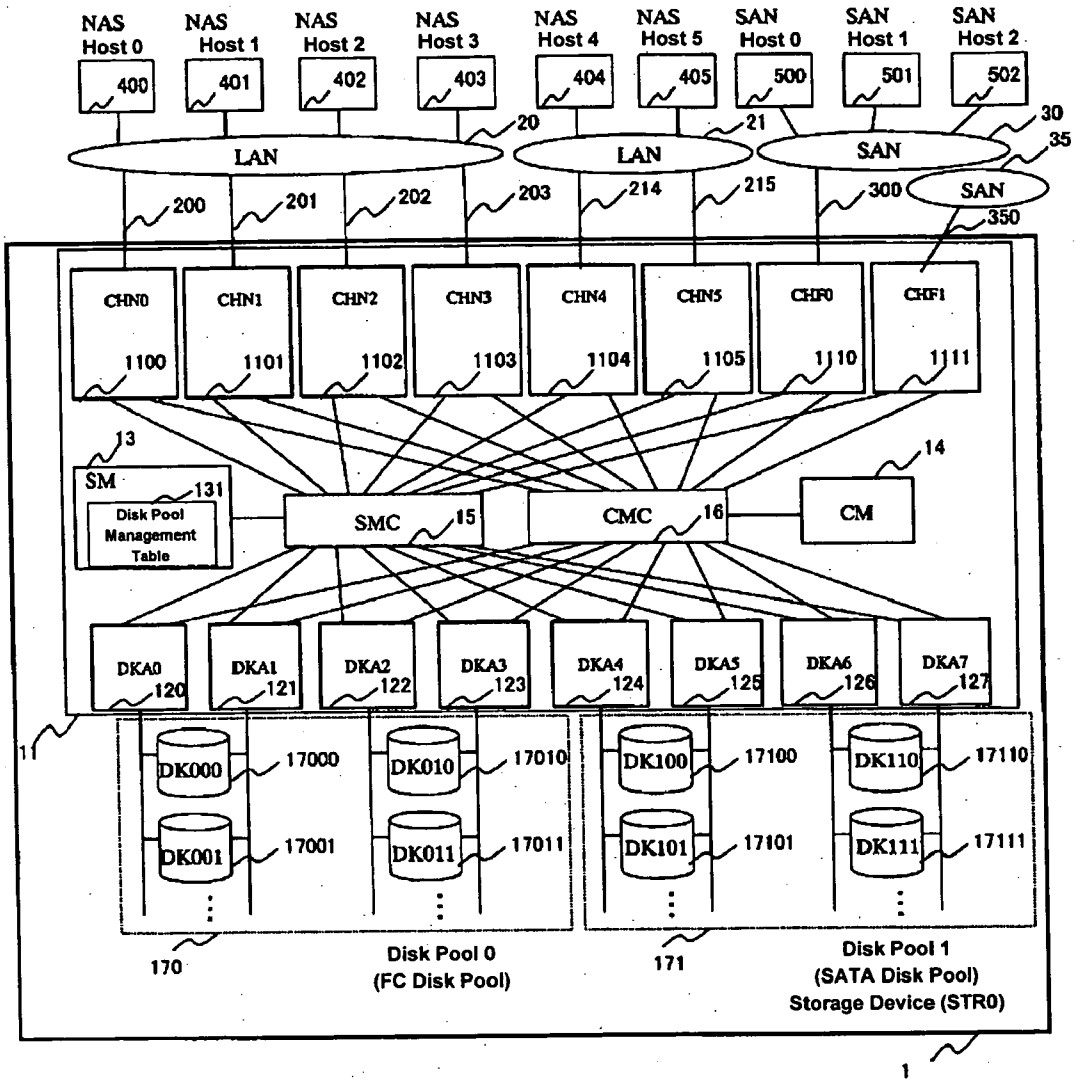


Fig. 1

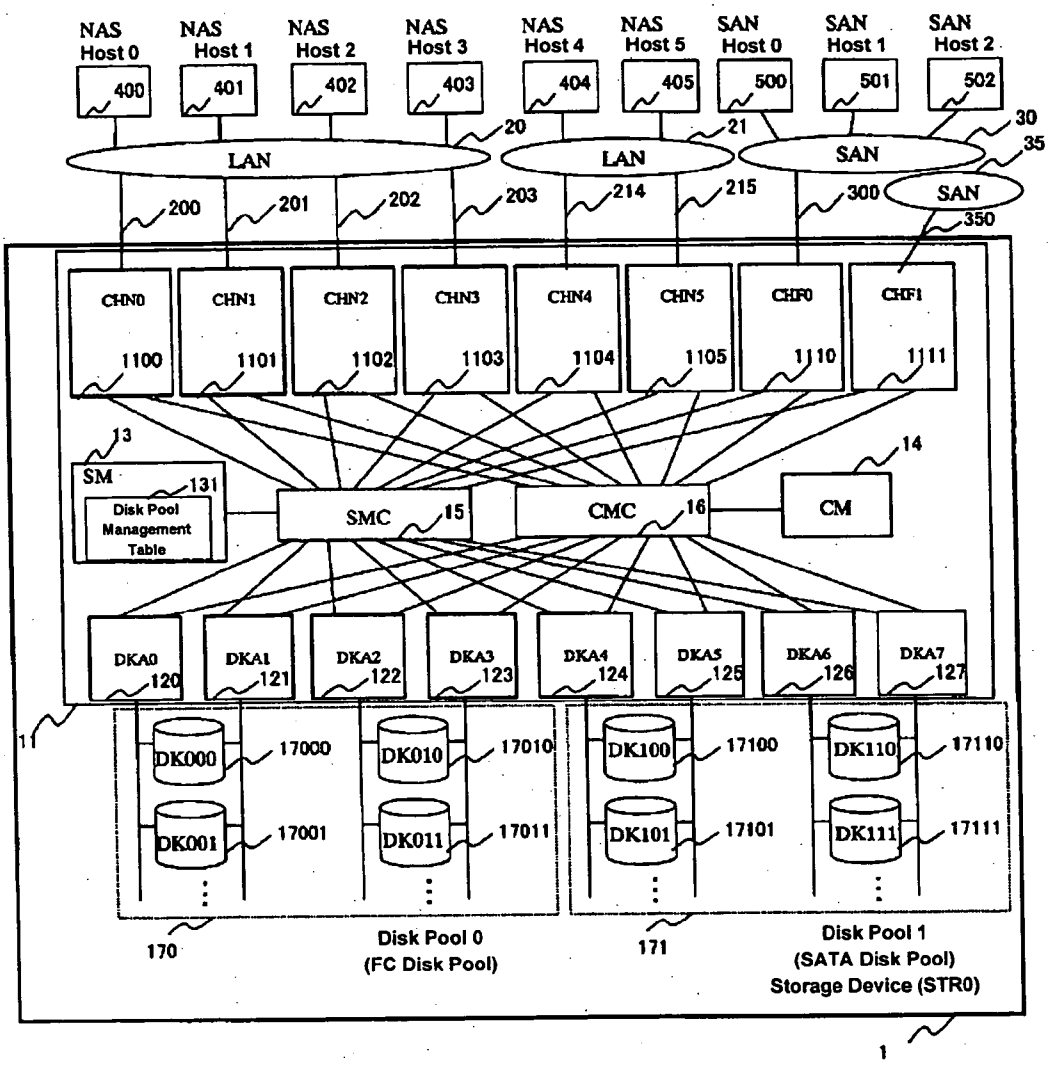

NAS Host 0
NAS Host 1
NAS Host 2
NAS Host 3
NAS Host 4
NAS Host 5
SAN Host 0
SAN Host 1
SAN Host 2
400
401
402
403
404
405
500
501
502
20
21
30
35
LAN
LAN
SAN
SAN
200
201
202
203
214
215
300
350
CHN0
CHN1
CHN2
CHN3
CHN4
CHN5
CHF0
CHF1
1100
1101
1102
1103
1104
1105
1110
1111
13
SM
131
Disk Pool Management Table
SMC
15
CMC
16
14
CM
DKA0
DKA1
DKA2
DKA3
DKA4
DKA5
DKA6
DKA7
120
121
122
123
124
125
126
127
11
DK000
DK001
DK010
DK011
DK100
DK101
DK110
DK111
17000
17001
17010
17011
17100
17101
17110
17111
170
171
Disk Pool 0
(FC Disk Pool)
Disk Pool 1
(SATA Disk Pool)
Storage Device (STR0)
1

Fig. 2

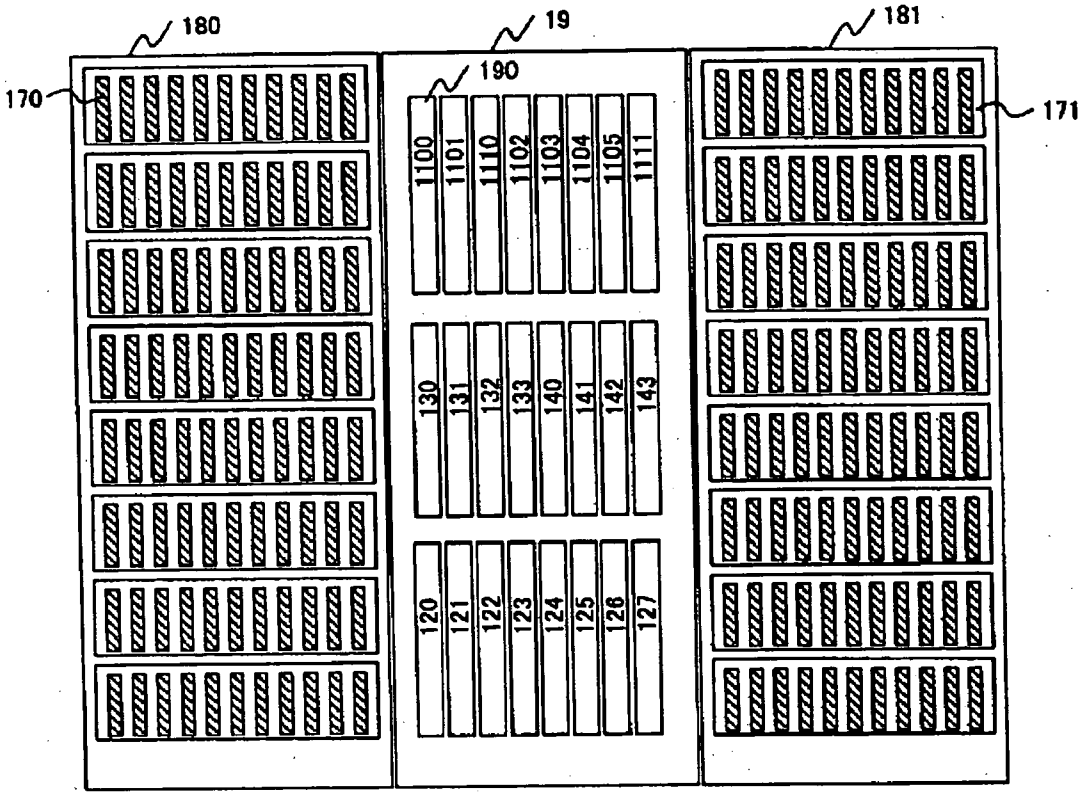

180
19
181
190
170
171
1100
1101
1110
1102
1103
1104
1105
1111
130
131
132
133
140
141
142
143
120
121
122
123
124
125
126
127

Fig. 3

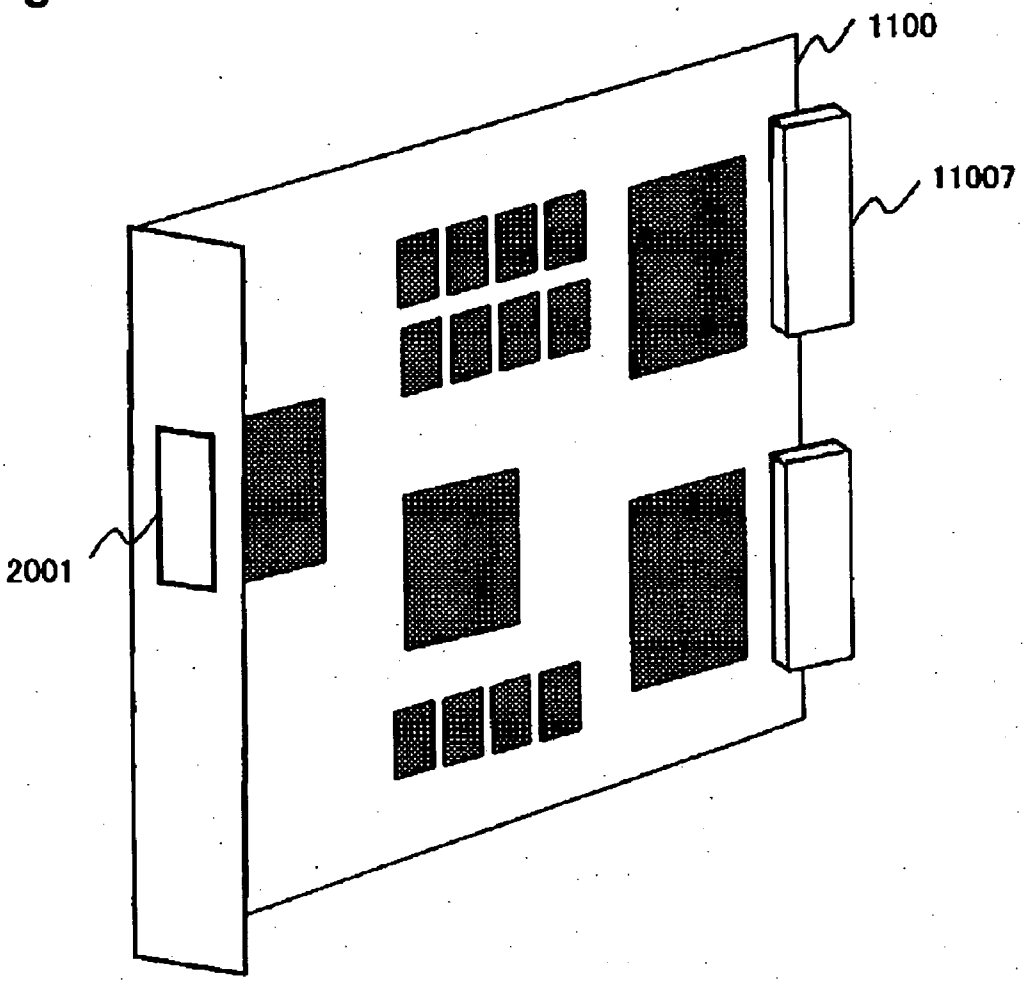


Fig. 4

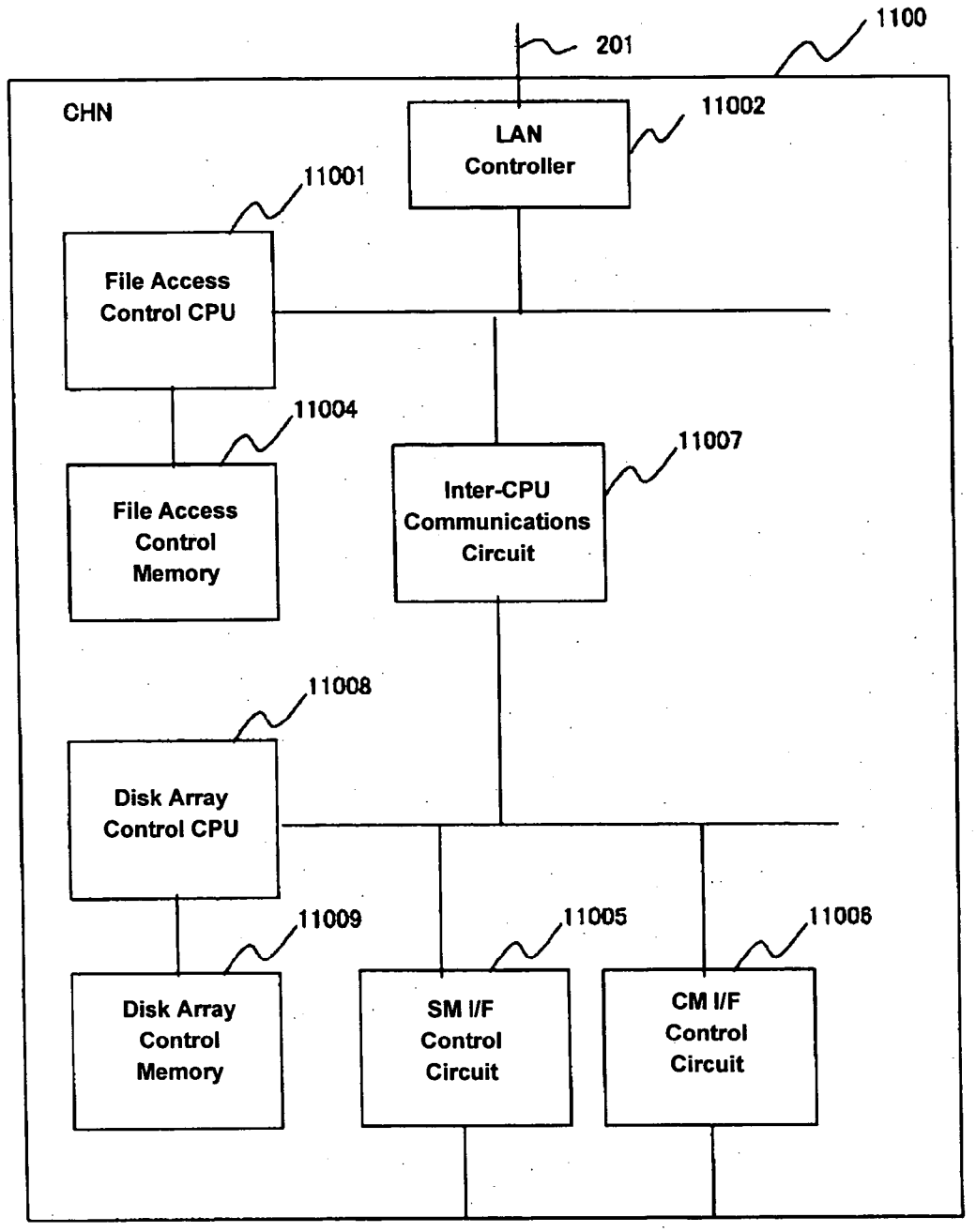

1100
201
CHN
LAN Controller
11002
11001
File Access Control CPU
11004
File Access Control Memory
Inter-CPU Communications Circuit
11007
11008
Disk Array Control CPU
11009
11005
11006
Disk Array Control Memory
SM I/F Control Circuit
CM I/F Control Circuit

Fig. 5

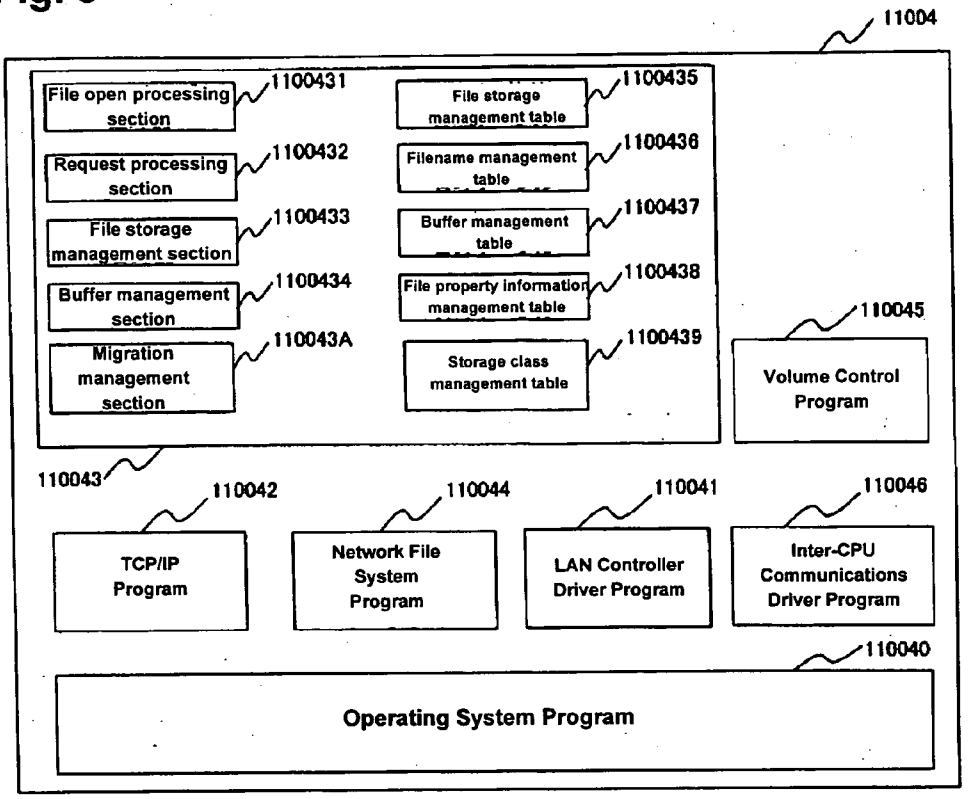

11004
File open processing section
1100431
Request processing section
1100432
File storage management section
1100433
Buffer management section
1100434
Migration management section
110043A
File storage management table
1100435
Filename management table
1100436
Buffer management table
1100437
File property information management table
1100438
Storage class management table
1100439
110045
Volume Control Program
110043
110042
TCP/IP Program
110044
Network File System Program
110041
LAN Controller Driver Program
110046
Inter-CPU Communications Driver Program
110040
Operating System Program

Fig. 6

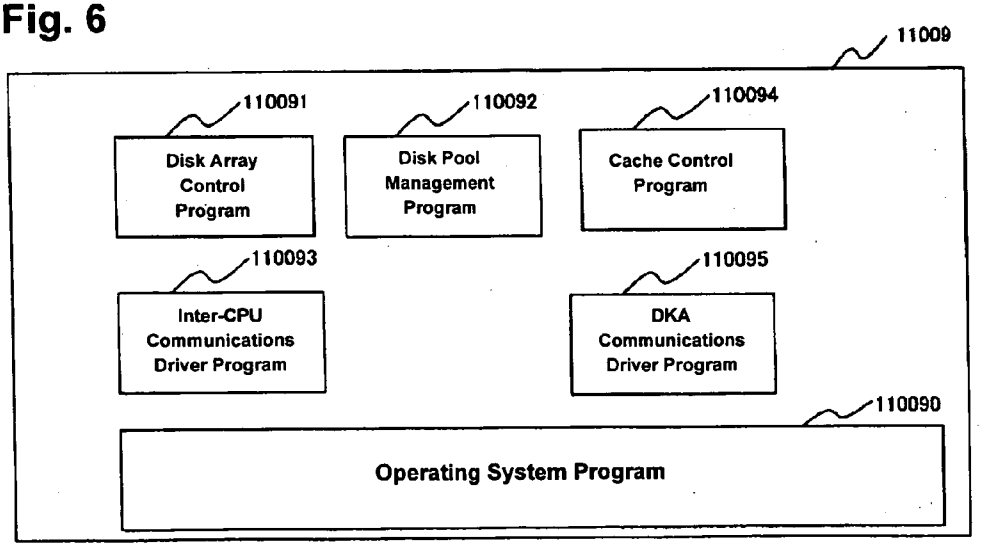

11009
110091
Disk Array Control Program
110092
Disk Pool Management Program
110094
Cache Control Program
110093
Inter-CPU Communications Driver Program
110095
DKA Communications Driver Program
110090
Operating System Program

Fig. 7

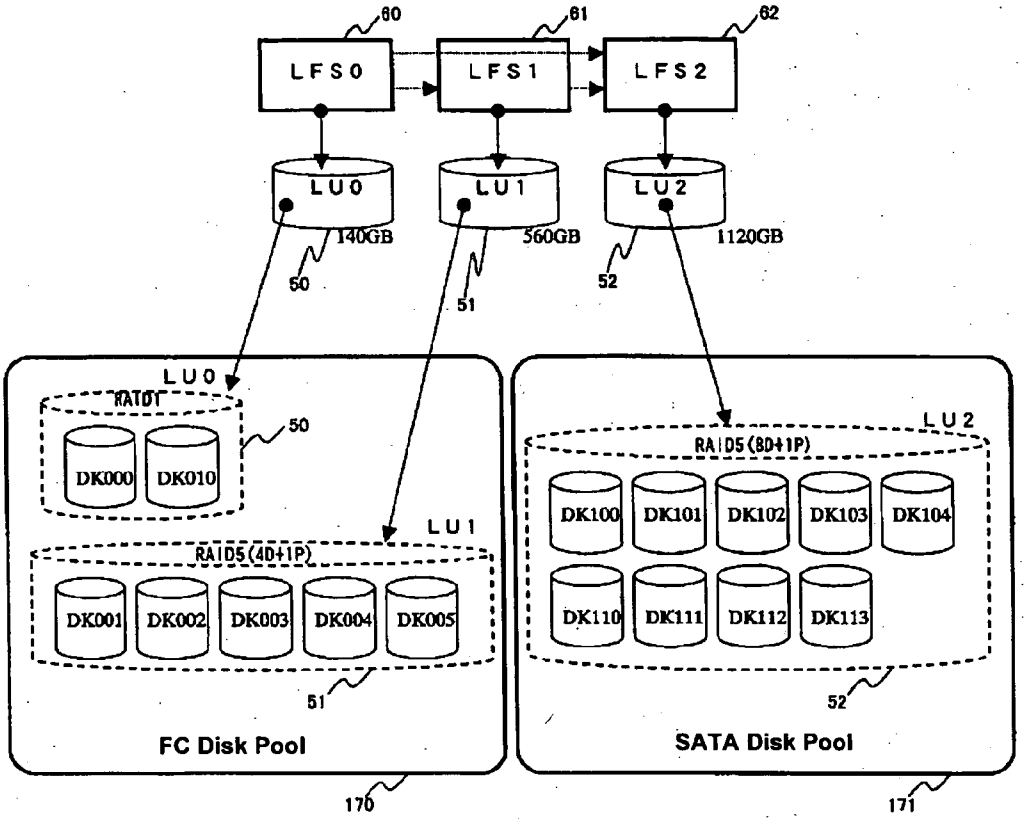

60
61
62
LFS0
LFS1
LFS2
LU0
LU1
LU2
140GB
560GB
1120GB
50
51
52
LU0
RAID1
50
DK000
DK010
LU1
RAID5(4D+1P)
DK001
DK002
DK003
DK004
DK005
51
FC Disk Pool
170
LU2
RAID5(8D+1P)
DK100
DK101
DK102
DK103
DK104
DK110
DK111
DK112
DK113
52
SATA Disk Pool
171

Fig. 8

Storage Class Management Table 1100451

Storage Class# (1100451a)	Storage Node# (1100451b)	Disk PooL# (1100451c)	LU# (1100451d)	LU Type (1100451e)	RAID Conf. (1100451f)	Usable Capacity (1100451g)	Used Capacity (1100451h)
OnLine Storage (Premium)	STR0	FC Pool	LU0	Local File	RAID1 1D+1P	140GB	53GB
OnLine Storage (Normal)	STR0	FC Pool	LU1	Local File	RAID5 4D+1P	560GB	125GB
NearLine Storage	STR0	SATA Pool	LU2	Local File	RAID5 8D+1P	1120GB	320GB
Archive Storage	STR1	SATA Pool	LU3	Remote File	RAID5 15D+1P	2100GB	480GB
Archive Storage	STR2	SATA Pool	LU4	Remote Block	RAID5 15D+1P	2100GB	480GB

Fig. 9

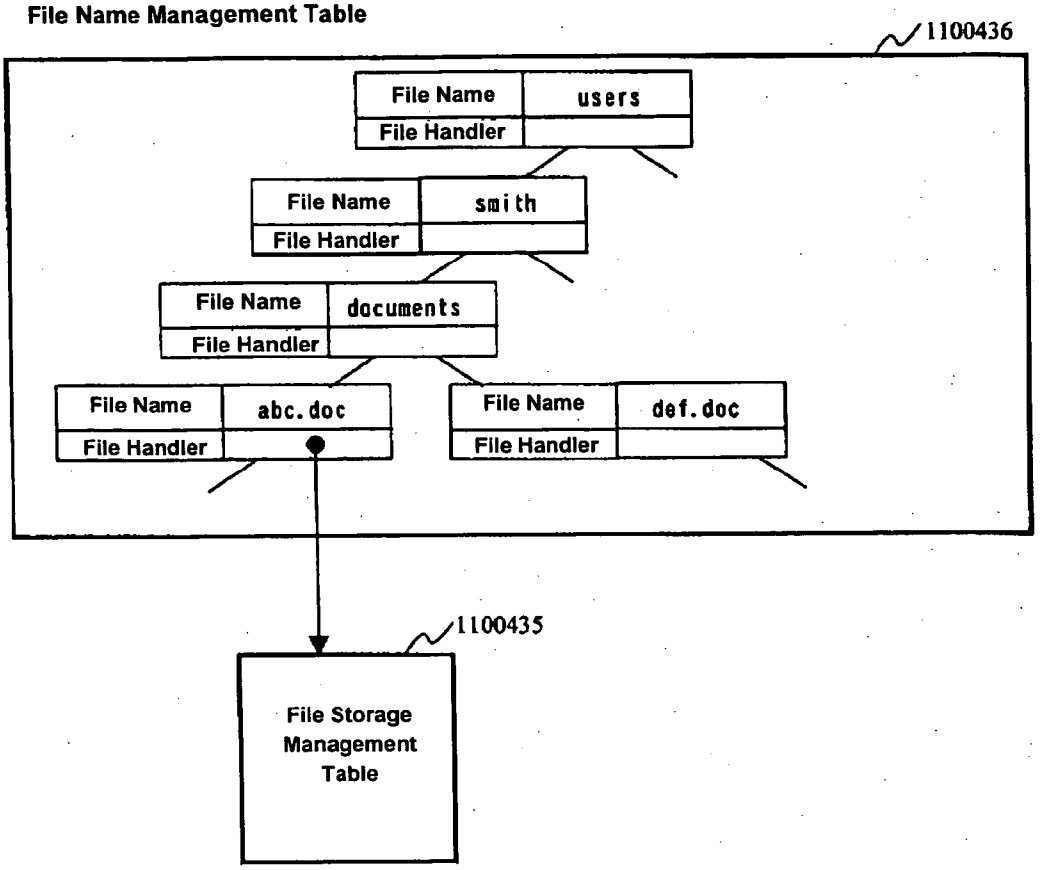

File Name Management Table
1100436
File Name
users
File Handler
File Name
smith
File Handler
File Name
documents
File Handler
File Name
abc.doc
File Handler
File Name
def.doc
File Handler
1100435
File Storage Management Table

Fig. 10

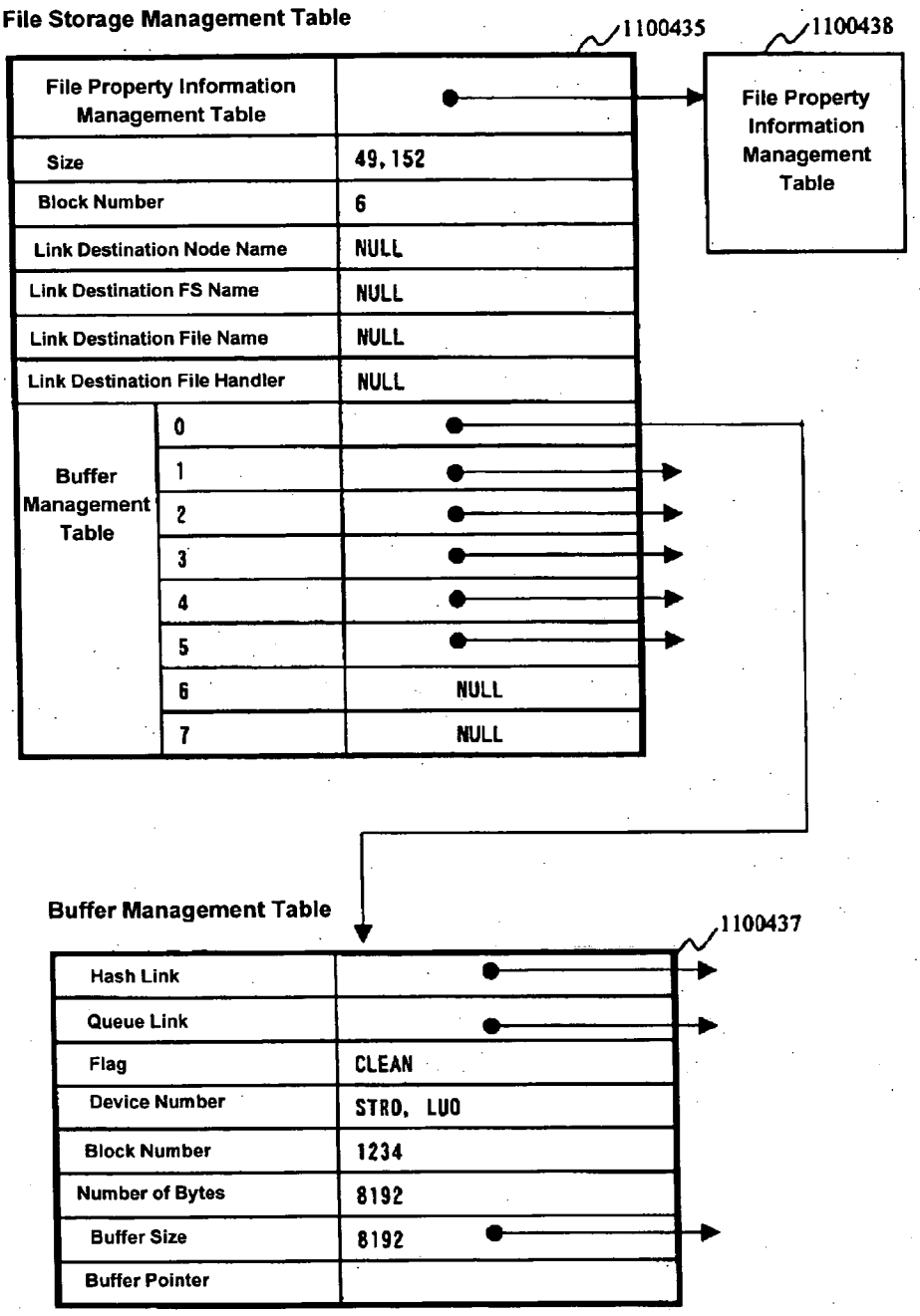

File Storage Management Table
1100435
1100438
File Property Information Management Table
File Property Information Management Table
Size
49, 152
Block Number
6
Link Destination Node Name
NULL
Link Destination FS Name
NULL
Link Destination File Name
NULL
Link Destination File Handler
NULL
Buffer Management Table
0
1
2
3
4
5
6
NULL
7
NULL
Buffer Management Table
1100437
Hash Link
Queue Link
Flag
CLEAN
Device Number
STRO, LUO
Block Number
1234
Number of Bytes
8192
Buffer Size
8192
Buffer Pointer

Fig. 11

File Property Information Management Table 1100438

Property Information Type	Category	Attribute	Content
Static Property Information	File Information	File Type	Document
		Application	XYZ Word
		Date Created	2002.11.25
		Owner	Smith
		Access Identifier	-rw-rw-rw-
	Policy	Initial Storage Class	Not designated
		Asset Value Type	Important
		Life Cycle Model	Model 1
		Migration Plan	Plan 1
Dynamic Property Information	Access Information	Time Stamp	2003.1.1
		Access Count	12
		Read Count	7
		Write Count	5
		Read Size	7,168
		Write Size	5,120
		Read Sequential Count	6
		Write Sequential Count	4
	Life Cycle stage Information	Current Life Cycle	Reference Stage
		Current Storage Class	NearLine

Fig. 12

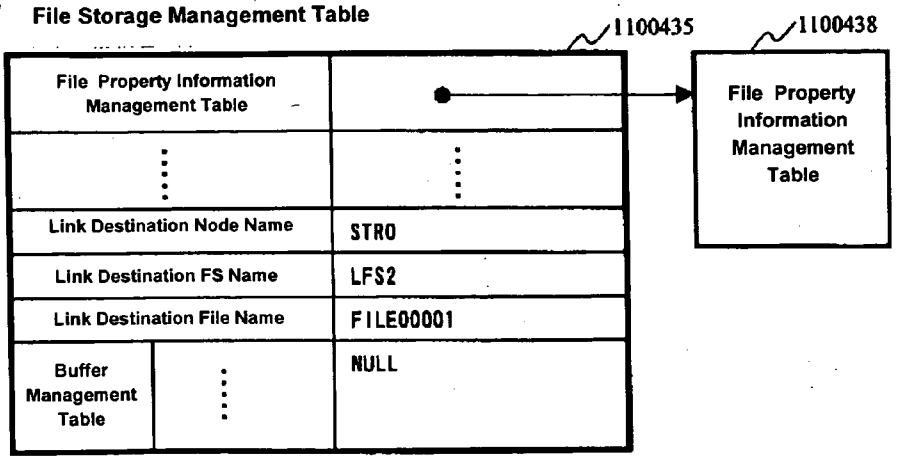


Fig. 13

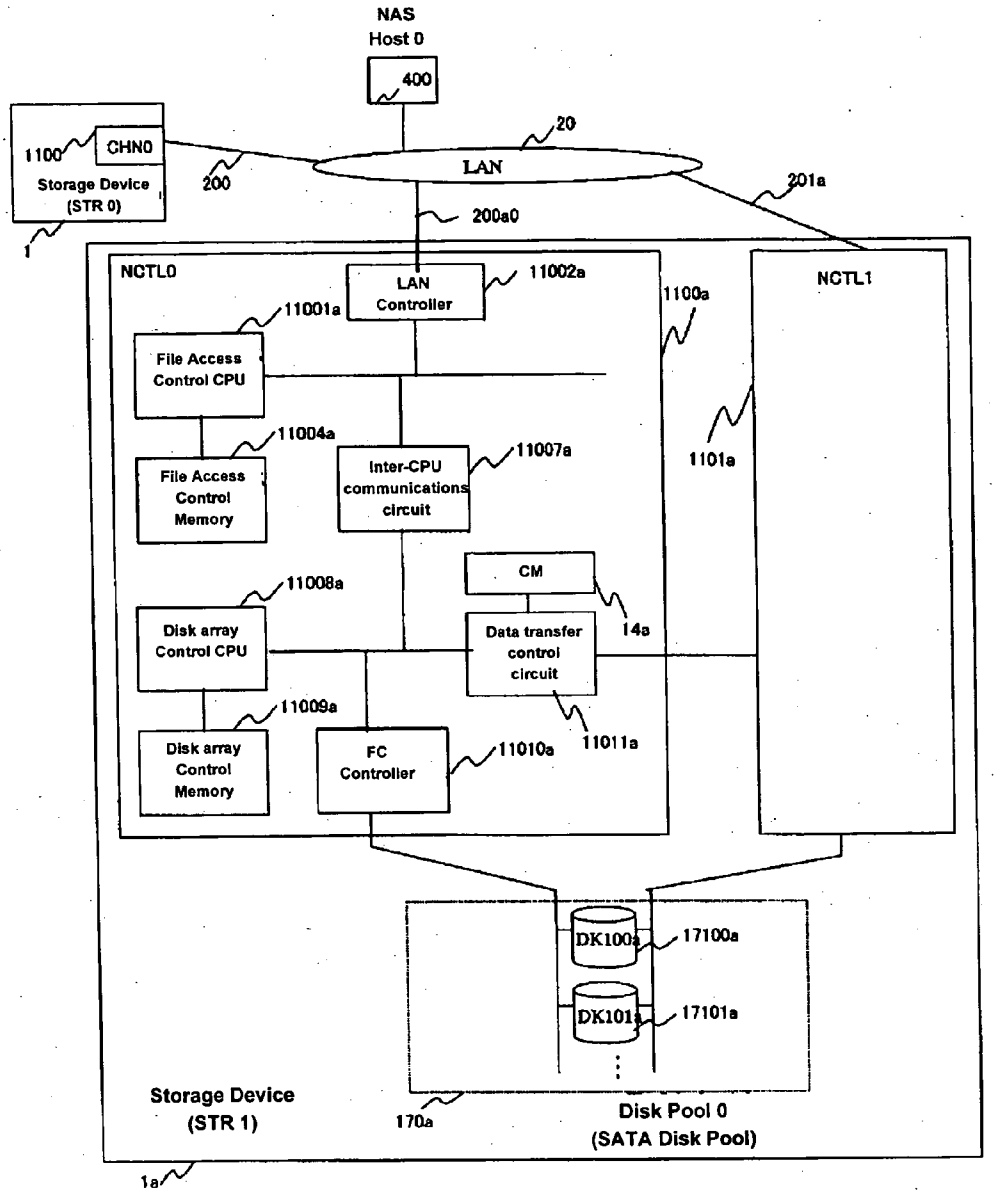

NAS
Host 0
400
20
LAN
1100
CHN0
Storage Device
(STR 0)
200
201a
200a0
1
NCTL0
LAN
Controller
11002a
11001a
1100a
NCTL1
File Access
Control CPU
11004a
File Access
Control
Memory
Inter-CPU
communications
circuit
11007a
1101a
CM
14a
11008a
Disk array
Control CPU
Data transfer
control
circuit
11009a
11011a
Disk array
Control
Memory
FC
Controller
11010a
DK100a
17100a
DK101a
17101a
Storage Device
(STR 1)
170a
Disk Pool 0
(SATA Disk Pool)
1a

Fig. 14

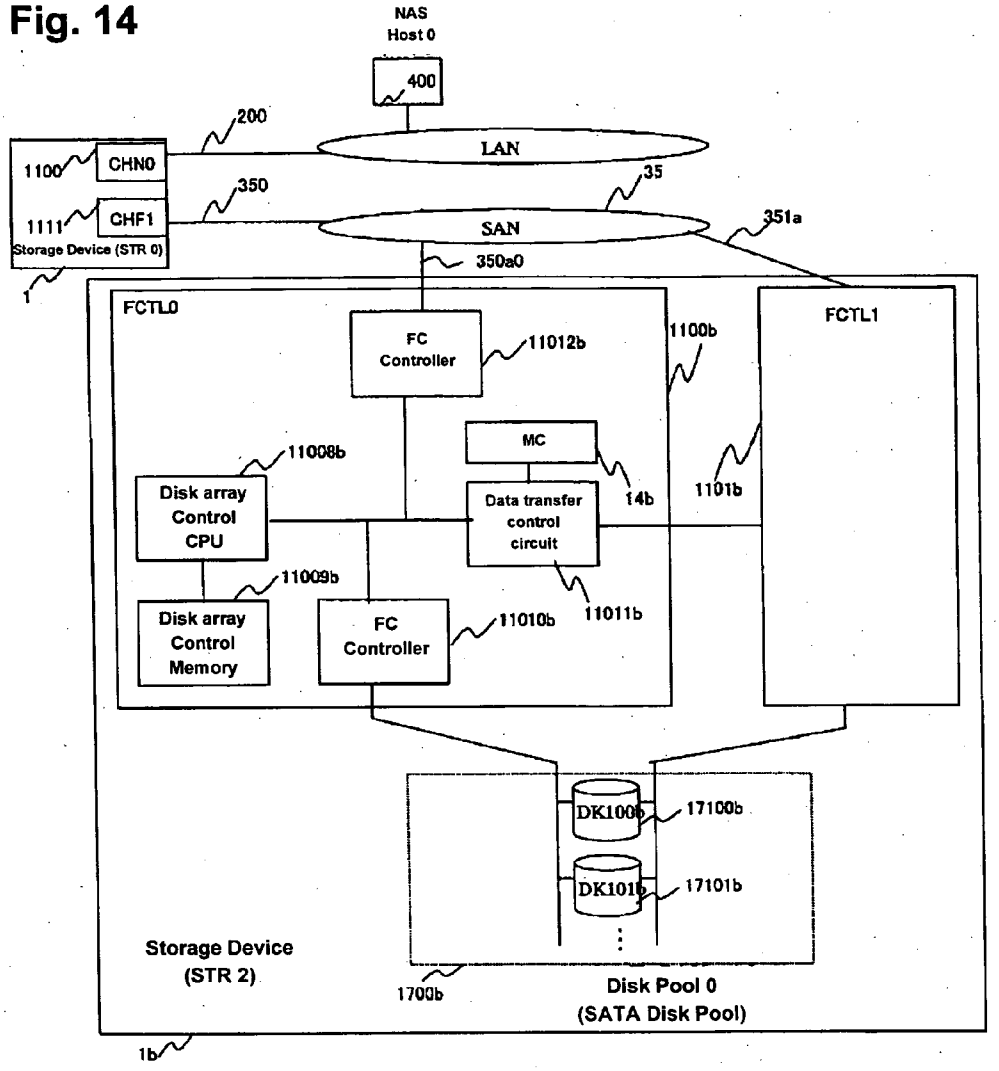

NAS
Host 0
400
200
LAN
1100
CHN0
350
35
1111
CHF1
SAN
351a
Storage Device (STR 0)
350a0
1
FCTL0
FC
Controller
11012b
1100b
FCTL1
MC
11008b
14b
1101b
Disk array
Control
CPU
Data transfer
control
circuit
11009b
11011b
Disk array
Control
Memory
FC
Controller
11010b
DK100b
17100b
DK101b
17101b
Storage Device
(STR 2)
1700b
Disk Pool 0
(SATA Disk Pool)
1b

Fig. 15

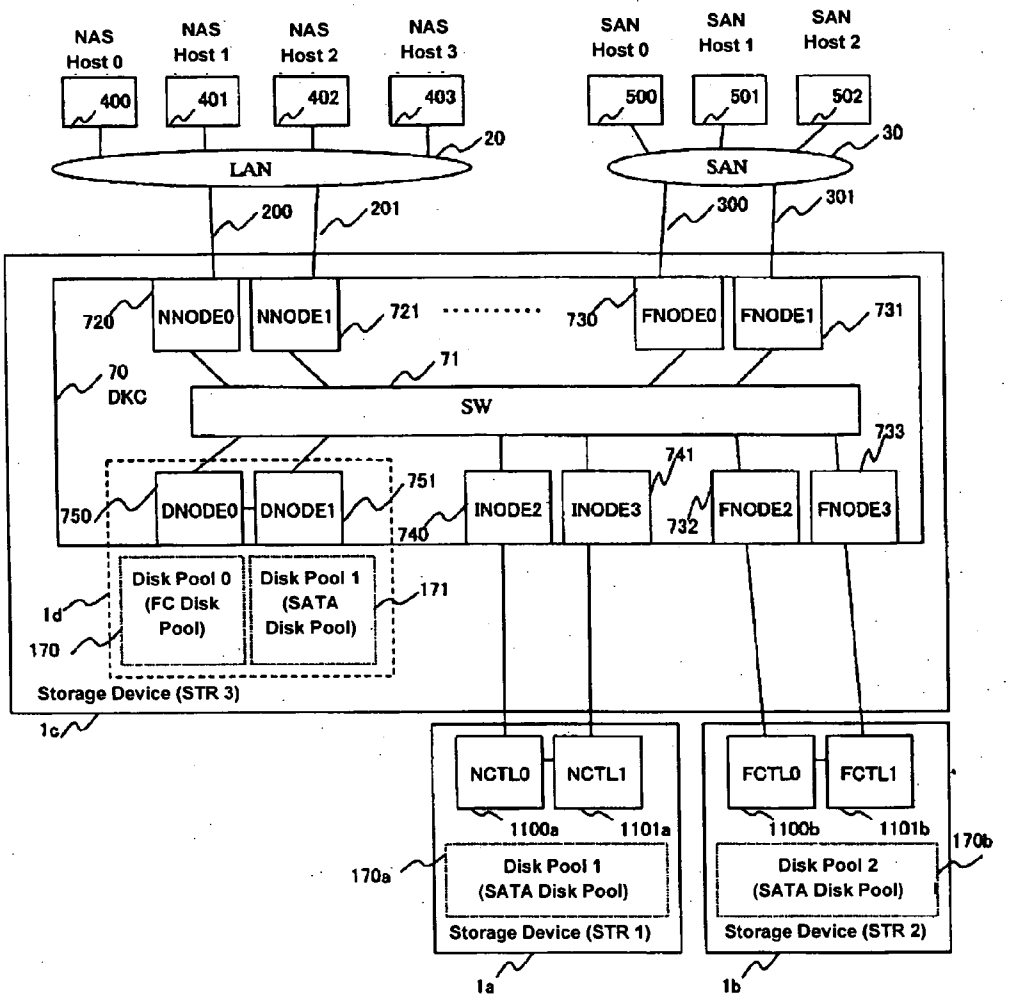

NAS Host 0
NAS Host 1
NAS Host 2
NAS Host 3
SAN Host 0
SAN Host 1
SAN Host 2
400
401
402
403
500
501
502
20
30
LAN
SAN
200
201
300
301
720
NNODE0
NNODE1
721
730
FNODE0
FNODE1
731
70
DKC
71
SW
733
741
750
DNODE0
DNODE1
751
740
INODE2
INODE3
732
FNODE2
FNODE3
Disk Pool 0 (FC Disk Pool)
Disk Pool 1 (SATA Disk Pool)
171
1d
170
Storage Device (STR 3)
1c
NCTL0
NCTL1
1100a
1101a
170a
Disk Pool 1 (SATA Disk Pool)
Storage Device (STR 1)
1a
FCTL0
FCTL1
1100b
1101b
170b
Disk Pool 2 (SATA Disk Pool)
Storage Device (STR 2)
1b

Fig. 16

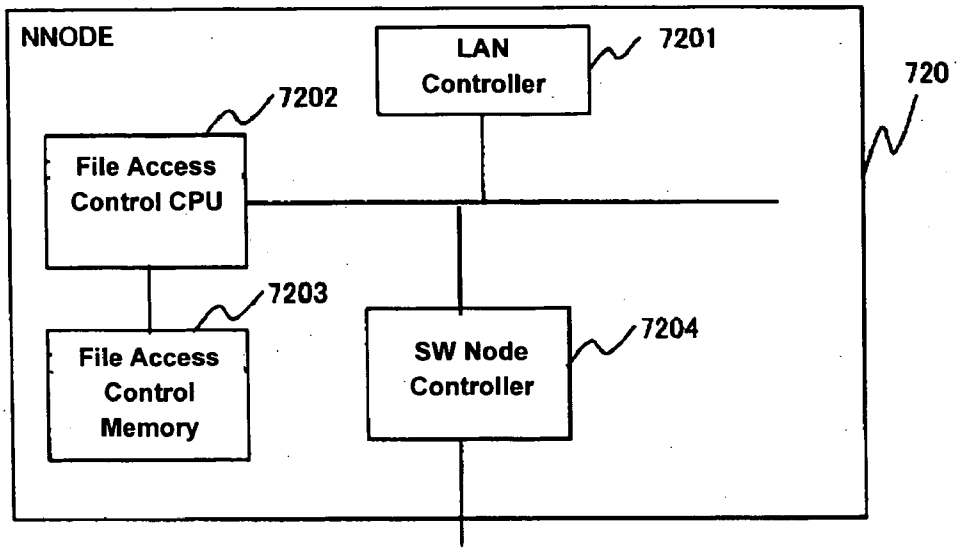

NNODE
LAN Controller
7201
720
7202
File Access Control CPU
7203
File Access Control Memory
SW Node Controller
7204

Fig. 17

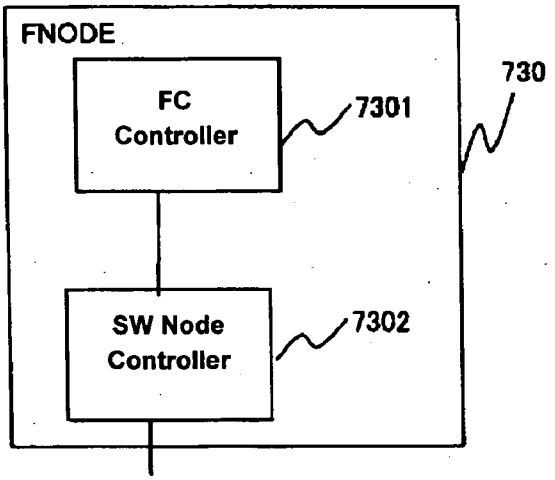

FNODE
730
FC Controller
7301
SW Node Controller
7302

Fig. 18

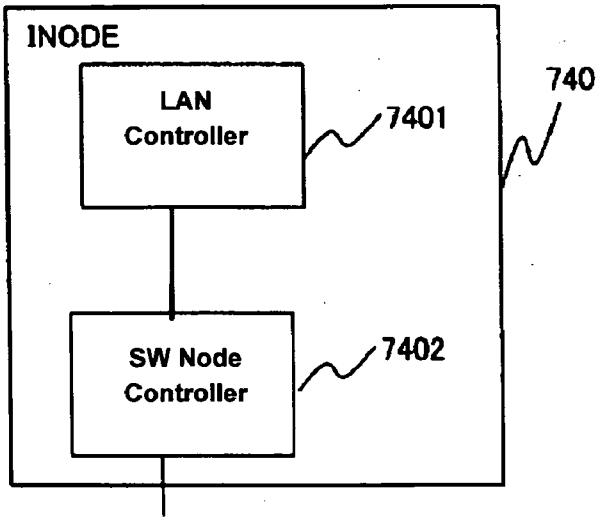
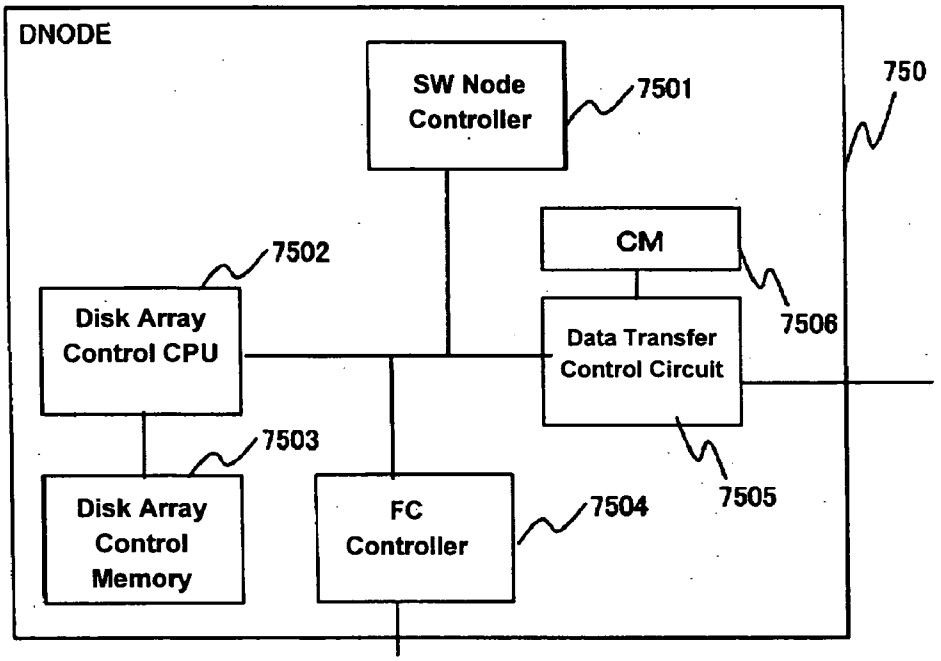

INODE
LAN Controller
7401
740
SW Node Controller
7402

Fig. 19


DNODE
SW Node Controller
7501
750
CM
7502
Disk Array Control CPU
7506
Data Transfer Control Circuit
7503
7505
Disk Array Control Memory
FC Controller
7504

STORAGE DEVICE

BACKGROUND OF THE INVENTION

[0001] 1. Field of the Invention

[0002] The present invention relates to a storage device used in a computer system.

[0003] 2. Related Background Art

[0004] In a conventional system, a high-speed storage device and a low-speed storage device may be connected to a computer called a hierarchical storage. In the system, files that are frequently used are stored in the high-speed storage device such as a magnetic disk device, while files that are not frequently used are stored in the inexpensive, low-speed storage device such as a tape device. Which files should be placed, i.e., stored, in which storage device is determined by using a table that manages access frequency of each file.

[0005] In another conventional system, a plurality of logical storage devices having different processing speeds and storage capacities are configured within a storage device that is connected to a computer and used. Such a system may be represented by disk array subsystems. In this system, the storage device manages as statistical information the frequency of accesses from the computer to data stored in the storage device, and, based on the statistical information, transfers data with high access frequency to logical storage devices with higher performance.

[0006] The first set of problems entailed in the prior art is that there is a high dependency on the computer connected to the storage device, that there is a limitation in the system configuration, and that it is difficult to simplify the system management.

[0007] In the conventional system described above, a hierarchical storage control is realized through software operating on the computer. The hierarchical storage control refers to a data storage control for controlling a plurality of storage regions having different processing speeds and storage capacities such that the storage regions can be changed according to the frequency of data usage. In other words, the hierarchical storage control refers to controlling to select, based on the property of data such as frequency of data usage, an appropriate storage region from among a plurality of storage regions having different properties in terms of processing speed and/or storage capacity, and to store the data in the storage region selected. However, when the system configuration is altered, such as when an old computer is replaced by a new computer, maintaining the system can be difficult due to such reasons as the system configuration of the new computer not being able to take over the software's control information.

[0008] Also in the conventional system described above, although a hierarchical storage control is implemented on a per-logical storage device basis, a technology for the storage device to recognize a data structure of data stored in the logical storage device or a technology for executing exclusive control are not disclosed. As a result, it would be difficult for a plurality of computers to share the same logical storage devices, and integrating storage devices used by a plurality of computers in order to reduce the management cost of the computer system would require imposing certain limitations on the configuration of the computer system, such as allocating a logical storage device for each computer.

[0009] The second problem is that optimal placement of data according to the life cycle or type of data is difficult.

[0010] According to the conventional technology, data that had high access frequency in the past is assumed to have high access frequency in the future as well, and the storage regions in which the data is stored are determined based on statistical information regarding data access frequency and on used capacity of storage regions that can be accessed at high-speed. The processing efficiency can be improved by increasing the probability with which data with high access frequency can reside in a storage device that can be accessed at high-speed. However, there are no technologies disclosed for determining storage regions in which to store data by taking into consideration differences in data properties that are dependent on the data's life cycle stage, i.e., the time elapsed since the corresponding file was generated, the type of application that generates and uses the data, and the type of data itself.

[0011] The third problem is that the effect of the hierarchical storage control is small.

[0012] Although the conventional system described above executes a hierarchical storage control by taking advantage of the difference in capacity and price between magnetic tapes and magnetic disks, the difference in capacity and price between magnetic tapes and magnetic disks have been growing smaller in recent years; consequently, the effect of cost optimization and cost reduction through the use of hierarchical storage control has also been growing smaller. Furthermore, due to the fact that the access speed to magnetic tapes is extremely slow compared to access speed to magnetic disks, it is difficult to use magnetic tapes as storage device for online access.

[0013] In the conventional system described above, a hierarchical storage control is executed by taking advantage of the difference in price and performance resulting from different RAID configurations of magnetic disks; however, since the price difference results only from the difference in the degree of redundancy in RAID configurations, the only cost reduction that can be hoped for is the cost reduction equivalent only to the difference in the degree of redundancy.

SUMMARY OF THE INVENTION

[0014] The present invention relates to a control method or a storage device that can execute a hierarchical storage control for file storage positions without being dependent on the OS or applications executed on a host computer.

[0015] The present invention also relates to a hierarchical storage control method for a plurality of computers to share files, or to provide a storage device that executes such a hierarchical storage control.

[0016] The present invention also relates to a control method or a storage device that can execute a hierarchical storage control according to file properties.

[0017] The present invention further relates to a hierarchical storage control method with high cost reduction effect, or to provide a storage device that executes such a hierarchical storage control.

[0018] In accordance with an embodiment of the present invention, a storage device comprises a plurality of storage regions having different properties, an interface control device that accepts from one or more computers access requests containing file identification information, and an interface control device for accessing storage regions that store data of the file designated by identification information, wherein the interface control device controls storage of file data in one of a plurality of storage regions according to the file property.

[0019] Other features and advantages of the invention will be apparent from the following detailed description, taken in conjunction with the accompanying drawings that illustrate, by way of example, various features of embodiments of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0020] **FIG. 1** is a diagram of an example of the configuration of a computer system in accordance with an embodiment of the present invention.

[0021] **FIG. 2** is a diagram of an example of the exterior appearance of a storage device.

[0022] **FIG. 3** is a diagram of an example of the exterior appearance of an adapter board.

[0023] **FIG. 4** is a diagram of an example of the configuration of a NAS channel adapter.

[0024] **FIG. 5** is a diagram of an example of programs stored in a file system control memory.

[0025] **FIG. 6** is a diagram of an example of programs stored in a disk array control memory.

[0026] **FIG. 7** is a diagram of an example of the relationship among disk pools, LUs and file systems.

[0027] **FIG. 8** is a diagram of an example of a storage class management table.

[0028] **FIG. 9** is a diagram of an example of a filename management table.

[0029] **FIG. 10** is a diagram of an example of a file storage management table and a buffer management table.

[0030] **FIG. 11** is a diagram of an example of a file property information management table.

[0031] **FIG. 12** is a diagram of an example of a file storage management table.

[0032] **FIG. 13** is a diagram of an example of the second configuration of a system in accordance with another embodiment of the present invention.

[0033] **FIG. 14** is a diagram of an example of the third configuration of the system in accordance with another embodiment the present invention.

[0034] **FIG. 15** is a diagram of an example of the fourth configuration of the system in accordance with another embodiment the present invention.

[0035] **FIG. 16** is a diagram of a configuration example of a NAS node.

[0036] **FIG. 17** is a diagram of a configuration example of a Fibre Channel node.

[0037] **FIG. 18** is a diagram of a configuration example of an IP node.

[0038] **FIG. 19** is a diagram of a configuration example of a disk array node.

PREFERRED EMBODIMENTS OF THE INVENTION

[0039] The following is a description of embodiments of the present invention. The following embodiments do not limit the present invention.

Embodiment 1

[0040] (1) Example of System Configuration (**FIG. 1**)

[0041] **FIG. 1** is a diagram indicating an example of a computer system including a storage device **1** (it is also called a storage system), to which the present invention is applied. In the following, x may be any integer.

[0042] The storage device **1** is a disk array system comprising a disk controller (hereinafter called "DKC") **11** and a plurality of magnetic disk devices (hereinafter simply called "disks") **170***x* and **171***x*. In the present embodiment, the storage device **1** is provided with two types of disks **170***x*and **171***x*. **170***x* are Fibre Channel (hereinafter called "FC") disks with FC-type interface, while **171***x* are serial AT attached (hereinafter called "SATA") disks with serial ATA-type (SATA-type) interface. A plurality of FC disks **170***x* makes up an FC disk pool **0** (**170**), while a plurality of SATA disks **171***x* makes up a SATA disk pool **1** (**171**). The disk pools will be described in detail later.

[0043] Next, the configuration of the DKC **11** of the storage device **1** will be described. The DKC **11** comprises one or more NAS channel adapters **110***x*, one or more Fibre Channel adapters **111***x*, a plurality of disk adapters **12***x*, a shared memory **13** (hereinafter called "SM"), a shared memory controller **15** (hereinafter called "SMC"), a cache memory **14** (hereinafter called "CM"), and a cache memory controller **16** (hereinafter called "CMC").

[0044] The NAS channel adapters (hereinafter called "CHN") **110***x* are interface control devices connected by file I/O interfaces to computers **40***x* (hereinafter called "NAS hosts"), which are connected to a local area network (hereinafter called "LAN") **20** or a LAN **21**.

[0045] The Fibre Channel adapters (hereinafter called "CHF") **111***x* are interface control devices connected by block I/O interfaces to computers (hereinafter called "SAN hosts") **50***x*, which are connected to a storage area network (hereinafter called "SAN") **30**. Hereinafter, CHN and CHF are collectively called channel adapters (hereinafter called "CH").

[0046] The disks **17***x* are connected to the disk adapters **12***x*. Each disk adapter (hereinafter called "DKA") **12***x* controls input and output to and from one or more disks **17***x* connected to itself.

[0047] The SMC **15** is connected to the CHN **110***x*, the CHF**111***x*, the DKA **12***x* and the SM **13**. The SMC **15** controls data transfer among the CHN **110***x*, the CHF **111***x*, the DKA **12***x* and the SM **13**. The CMC **16** is connected to the CHN **110***x*, the CHF**111***x*, the DKA **12***x* and the CM **14**.

The CMC **16** controls data transfer among the CHN **110***x*, the CHF**111***x*, the DKA **12***x* and the CM **14**.

[**0048**] The SM **13** stores a disk pool management table **131**. The disk pool management table **131** is information that is used to manage the configuration of the disk pools.

[**0049**] The LANs **20** and **21** connect the CHNs **110***x* to the NAS hosts **40***x*. Generally, Ethernet ® is used for LAN. The SAN **30** connects the CHFs **111***x* to the SAN hosts **50***x*. Generally, Fibre Channel is used for SAN. However, an IP network can be used as the SAN, such that iSCSI, by which SCSI commands according to SCSI protocol are encapsulated into IP packets for sending and receiving, is used among equipment connected to the SAN. The SAN **35** according to the present embodiment is a dedicated SAN for connecting the storage device **1** and no SAN hosts are connected to the SAN **35**.

[**0050**] In the storage device **1**, all CHs can access the CM **14**, the SM **13**, any DKAs **12***x* and any disks **17***x*, via the CMC **16** or the SMC **15**.

[**0051**] The storage device **1** shown in **FIG. 1** has both the SAN interfaces (CHFs **111***x*) for connecting to the SAN hosts **50***x* and the NAS interfaces (CHNs **111***x*) for connecting to the NAS hosts **40***x*, but the present embodiment can be implemented even if the storage device **1** has only the NAS interfaces.

[**0052**] (2) Example of Exterior Appearance of Storage Device (**FIG. 2**)

[**0053**] **FIG. 2** is a diagram of an example of the exterior appearance of the storage device **1**.

[**0054**] A DKC unit **19** stores the CHNs **111***x*, the CHFs **110***x*, the DKAs **12***x*, the SM **13** and the CM **14**, which are components of the DKC **11**. The SM **13** actually comprises a plurality of controller boards **13***x*. The CM **14** also comprises a plurality of cache boards **14**x. Users of the storage device **1** can increase or decrease the number of such boards in order to configure the storage device **1** with the CM **14** and the SM **13** having the desired storage capacity. Disk units (hereinafter called "DKU") **180** and DKUs **181** store the disk pool **170** and the disk pool **171**, respectively.

[**0055**] Adapter boards built-in with the CHNs **110***x*, the CHFs **111***x*, the DKAs **12***x*, the controller boards **13***x* and the cache boards **14***x* are stored in slots **190**. According to the present embodiment, the shape of the slots **190**, the size of the adapter boards and the shape of connectors are made uniform regardless of the type of adapter boards or the type of interface, which maintains compatibility among various types of boards. As a result, in the DKC unit **19**, any adapter board can be mounted into any slot **190** regardless of the type of the adapter board or the type of the interface. Furthermore, users of the storage device **1** can freely select the number of adapter boards for the CHNs **110***x* and the CHFs **111***x* in order to mount the number of the CHNs **110***x* and the CHFs **111***x* selected into the slots **190** of the DKC unit **19**.

[**0056**] (3) Example of Exterior Configuration of Adapter Board (hereinafter called "NAS board") with the CHN **110**X Built-in (**FIG. 3**)

[**0057**] **FIG. 3** is a diagram of an example of the exterior configuration of a NAS board. A connector **11007** is connected to a connector of the DKC unit **19**. An interface connector **2001** is Ethernet ®-compatible and can be connected to Ethernet ®.

[**0058**] According to the present embodiment, due to the fact that the shape of the connector on adapter boards is uniform regardless of the type of the adapter board as described earlier, the adapter boards with built-in CHNs **110***x* and the adapter boards with built-in CHFs **111***x* have connectors of the same shape. On the adapter boards with the built-in CHFs **111***x*, the interface connector **2001** is Fibre Channel-compatible and configured to be connected to Fibre Channel.

[**0059**] (4) Example of Configuration of NAS Board (or CHN) (**FIG. 4**)

[**0060**] **FIG. 4** is a diagram of an example of the configuration of the CHN **110***x*. A file access control CPU **11001** is a processor for controlling file access. A LAN controller **11002** is connected to the LAN **20** via the interface connector **2001** and controls sending and receiving of data to and from the LAN **20**. A file access control memory **11004** is connected to the file access control CPU **11001**. The file access control memory **11004** stores programs executed by the file access control CPU **11001** and control data.

[**0061**] A disk array control CPU **11008** is a processor for controlling a disk array. The disk array refers to a storage device consisting of a plurality of disks. Disk arrays in which at least one of a plurality of disks stores redundant data to provide fault tolerance are called RAIDs. RAIDs are described later. A disk array control memory **11009** is connected to the disk array control CPU **11008** and stores programs executed by the disk array control CPU **11009** and control data. An SM I/F control circuit **11005** is a circuit for controlling access from the CHNs **110***x* to the SM **13**. A CM I/F control circuit **11006** is a circuit for controlling access from the CHNs **110***x* to the CM **14**. An inter-CPU communications circuit **11007** is a communications circuit used when the file access control CPU **11001** communicates with the disk array control CPU **11008** in order to access disks.

[**0062**] The present embodiment indicates an example of an asymmetrical multiprocessor configuration in which two processors, the file access control CPU **11001** and the disk array control CPU **11208**, are mounted on each CHN **110***x*; however, each CHN **110***x* can be configured by mounting a single processor that executes both the file access control and the disk array control, or as a symmetrical multiprocessor configuration in which two or more processors are mounted as equivalents to execute the file access control and the disk array control.

[**0063**] The configuration of each CHF**111***x* is the configuration shown in **FIG. 4**, except that components shown in top half of **FIG. 4**, namely the LAN controller **11002**, the file access control CPU **11001**, the file access control memory **11004** and the inter-CPU communications circuit **11007**, are replaced by a Fibre Channel controller.

[**0064**] (5) Example of Programs Stored in File Access Control Memory (**FIG. 5**)

[**0065**] **FIG. 5** is a diagram of an example of programs and control data stored in the file access control memory **11004** of the CHN **110***x*. An operating system program **110040** is used for the management of programs as a whole and for

input/output control. A LAN controller driver program **110041** is used for the control of the LAN controller **11002**. A TCP/IP program **110042** is used for the control of TCP/IP, which is the communications protocol for LAN. A file system program **110043** is used for managing files stored in the storage device **1**. A network file system program **110044** is used for controlling NFS and/or CIFS, which are protocols for providing files stored in the storage device **1** to the NAS hosts **40***x*. A volume control program **110045** is used for controlling the configuration of each logical volume by combining a plurality of logical disk units (hereinafter called "LU"), each of which is a unit of storage region set within the disk pools **17***x*. An inter-CPU communications driver program **110046** is used for controlling the inter-CPU communications circuit **11007**, which is used for communication between the file access control CPU **11001** and the disk array control CPU **11008**.

[**0066**] The file system program **110043** includes the following:

[**0067**] 1) a file open processing section **1100431** for executing a file open processing when using a file;

[**0068**] 2) a request processing section **1100432** for executing a processing according to a file access request when a file access request is received;

[**0069**] 3) a file storage management section **1100433** for dividing each file into blocks, determining the storage position on a disk for each block, and managing the storage position of each block;

[**0070**] 4) a buffer management section **1100434** for managing correlation between each block and a buffer formed in the memory;

[**0071**] 5) a file storage management table **1100435** for managing addresses of storage regions on disks that store blocks that make up each file;

[**0072**] 6) a filename management table **1100436** for managing filenames of open files and file handlers used to access the file storage management table **1100435** of each file;

[**0073**] 7) a buffer management table **1100437** for managing buffer addresses indicating storage regions within buffers corresponding to blocks that make up a file;

[**0074**] 8) a file property information management table **1100438** for storing file static properties, such as the file type, the application that generated the file, the intent of the file generator, and file dynamic properties, such as the value of the file that varies according to the file's life cycle stage and the file's access properties;

[**0075**] 9) a migration management section **110043**A used when executing a processing to migrate files between LUs; and

[**0076**] 10) a storage class management table **1100439** that registers for each LU in the storage pool a storage class, described later, and identification information of the storage device in which an LU resides.

[**0077**] (6) Configuration of Disk Array Control Memory (**FIG. 6**)

[**0078**] **FIG. 6** is a diagram of an example of programs stored in the disk array control memory **11009**. An operating system program **110090** is used for managing programs as a whole and for controlling input/output. A disk array control program **110091** is used for constructing LUs within the disk pools **17***x* and for processing access requests from the file access control CPU **11001**. A disk pool management program **110092** is used for managing the configuration of the disk pools **17***x* by using information in the disk pool management table **131** stored in the SM **13**. An inter-CPU communications driver program **110093** is used for controlling the inter-CPU communications circuit **11007**, which is used for communication between the file access control CPU **11001** and the disk array control CPU **11008**. A cache control program **110094** is used for managing data stored in the CM **14** and for controlling cache hit/miss judgments. A DKA communications driver program **110095** is used when accessing an LU in order to communicate with the DKAs **12***x*, which control the disks **170***x* and **171***x* that make up the LU.

[**0079**] (7) Configuration of Disk Pools (**FIG. 7**)

[**0080**] **FIG. 7** is a diagram of an example of the configuration of the disk pools.

[**0081**] In the FC disk pool **170** are set two LUs, LU**0** (**50**) and LU**1** (**51**). The LU**0** (**50**) comprises two FC disks, DK**000** and DK**010**, where the DK**000** and the DK**010** make up RAID **1**. The LU**1** (**51**) consists of 5FC disks, DK**001**, DK**002**, DK**003**, DK**004**, and DK**005**, where the five FC disks make up a 4D +1P configuration RAID **5**. The RAID **1** and the RAID **5** refer to data placement methods in a disk array and are discussed in detail in "A Case for Redundant Arrays of Inexpensive Disks (RAID)" by D. Patterson, et al., ACM SIGMOD Conference Proceedings, 1988, pp. 109-116. In the LU**0** with the RAID **1** configuration, the two FC disks DK**000** and DK**010** have a mirror relationship with each other. In the meantime, the LU**1** having the RAID **5** configuration consists of one or more disks that store data stripes, which store data of files accessed from host computers, and one or more disks that store parity stripes, which are used to retrieve data stored in the data stripes. The LU**1** has the 4D +1P configuration RAID **5**, which indicates a RAID **5** consisting of four data stripes and one parity stripe. Similar representations will be used hereinafter to indicate the number of data stripes and the number of parity stripes in LUs having the RAID **5** configuration.

[**0082**] LU**2** (**52**) is established in the SATA disk pool **171**. The LU**2** (**52**) consists of nine SATA disks, DK**100**, DK**101**, DK**102**, DK**103**, DK**104**, DK**110**, DK**1**, DK**112** and DK**113**, where the nine SATA disks make up an 8D +1P configuration RAID **5**.

[**0083**] When the capacity of each disk is 140 GB, the LU**0** (**50**) has 140 GB, the LU**1** (**51**) has 560 GB, and the LU**2** (**52**) has 1120 GB in usable storage capacity.

[**0084**] Independent local file systems LFS**0** (**60**), LFS**1** (**61**) and LFS**2** (**62**) are established and constructed for the LU**0** (**50**), the LU**1** (**51**) and the LU**2** (**52**), respectively.

[0085] (8) Storage Class Management Table (**FIG. 8**)

[0086] **FIG. 8** is an example of the configuration of a storage class management table **1100451** stored in the file access control memory **11004** of each CHN **110***x*. The storage class management table **1100451** is generated by the file access control CPU **11001**'s executing the file system program **110043** and referring to information stored in the disk pool management table **131** of the SM **13**.

[0087] Although the disk pool management table **131** is not shown, the disk pool management table **131** is stored in the SM **13** and contains information similar to the information in the storage class management table **1100451** for all CHs. In other words, of the information in the disk pool management table **131**, the storage class management table **1100451** stored in the file access control memory **11004** of each CHN **110***x* contains information regarding LUs used by the CHN **110***x*, but rearranged with the storage class as a key.

[0088] The following is a description of the configuration of the storage class management table **1100451**. A storage class entry (**1100451***a*) stores information indicating storage class. A storage node # entry (**1100451***b*) stores an identification number (called a "storage node number") of the storage device that makes up each storage class. A disk pool # entry (**1100451***c*) stores a disk pool number that makes up each storage class. An LU# entry (**1100451***d*) stores an LU number set for each disk pool. In an LU type entry (**1100451***e*), information stored indicates whether the corresponding LU is set internally (local) or externally (remote) to the given storage device and whether a file system is set in the LU. In other words, if the LU is within the storage device, "Local" is registered in the LU type entry, while "Remote" is registered if the LU is in a different storage device; if a file system is constructed in that LU, "File" is registered in the LU type entry, while "Block" is registered if no file systems are constructed in the LU. In a RAID Conf. entry (**1100451***f*), information stored indicates the RAID level of the disk array that makes up each LU and the structure of the corresponding disk array, such as data record and parity record number within a parity group. In a usable capacity entry (**1100451***g*) and a used capacity entry (**1100451***h*), information that indicates the total storage capacity of the given LU and information that indicates the storage capacity being used, respectively, are stored.

[0089] A storage class is a hierarchical attribute provided for each storage region based on the usage of data storage; according to the present embodiment, three attributes of OnLine Storage, NearLine Storage and Archive Storage are defined. In addition, sub-attributes of Premium and Normal are defined for the OnLine Storage. The OnLine Storage is an attribute set for LUs suitable for storing data of files that are frequently accessed, such as files being accessed online and files being generated. Premium indicates an attribute set for LUs suitable for storing data especially requiring fast response. The NearLine Storage is an attribute set for LUs suitable for storing data of files that are not frequently used but are occasionally accessed. The Archive Storage is an attribute set for LUs suitable for storing data of files that are hardly ever accessed and are maintained for long-term storage.

[0090] **FIG. 8** indicates that there are the LU**0** (**50**) of the OnLine Storage (Premium) class and the LU**1** (**51**) of the OnLine Storage (Normal) class in the FC disk pool **170** of the storage device **1** (called "STR**0**"). Further, in the SATA disk pool **171** of the storage device **1** (STR**0**) is the LU**2** (**52**) of the NearLine Storage class. Moreover, in a different storage device (STR**1**) is an LU**3** (**53**) of the Archive Storage class in a SATA disk pool. An example of constructing disk pools in different storage devices is described later.

[0091] (9) Filename Management Table (**FIG. 9**)

[0092] **FIG. 9** shows an example of the filename management table **1100436** that is stored in the file access control memory **11004**. The filename management table **1100436** is a table prepared for each file system, where filenames and file handlers are stored in a tree structure for easy searchability. When a file is accessed by one of the NAS hosts **40***x*, the filename of the file is included in an access request received by the CHN **110***x* from the NAS host **40***x*. The CHN **110***x* uses the filename to search the filename management table **11004** and obtains the file handler that corresponds to the filename, which enables the CHN **110***x* to refer to the file storage management table **1100435** that corresponds to the file handler.

[0093] Each filename management table **1100436** is stored in the LU in which the file system that corresponds to the filename management table **1100436** is constructed, and is read to the file access control memory **11004** when necessary and used by the file access control CPU **11001**.

[0094] (10) File Storage Management Table (**FIG. 10**)

[0095] **FIG. 10** is a diagram of an example of the file storage management table **1100435** and the buffer management table **1100437**. The file storage management table **1100435** is provided in the file access control memory **11004** for each file and is a table that manages file storage addresses. The file storage management table **1100435** can be referred to by designating a file handler that represents a file.

[0096] A file property information management table entry stores a pointer for referring to the file property information management table **1100438** for the corresponding file. A size indicates the size of the file in units of bytes. A number of blocks indicates the number of logical blocks used in managing the file, which is done by dividing the file into blocks called logical blocks. Each logical block that stores the file also stores a pointer to the buffer management table **1100437** that corresponds to the logical block.

[0097] There is one buffer management table **1100437** for each logical block, and each buffer management table **1100437** contains the following. A hash link entry stores a link pointer to a hash table for quickly determining whether a buffer is valid. A queue link entry stores a link pointer for forming a queue. A flag entry stores a flag that indicates the status of the corresponding buffer, i.e. whether valid data is stored in the buffer, whether the buffer is being used, whether the content of the buffer is unreflected on the disk. An equipment number entry stores an identifier of the storage device and an identifier of the LU in which the corresponding logical block is stored. A block number entry stores a disk address number that indicates the storage position of the logical block within the storage device indicated by the equipment number. A number of bytes entry stores the number of bytes of valid data stored in the logical block. A buffer size entry stores the size of the buffer in units

of bytes. A buffer pointer entry stores a pointer to the corresponding physical buffer memory.

[0098] The file storage management table **1100435** is stored in the LU that stores the corresponding file and is read to the memory when necessary for use.

[0099] (11) File Property Information Management Table (**FIG. 11**)

[0100] **FIG. 11** is an example of the file property information management table **1100438** stored in the file access control memory **11004**. The file property information management table **1100438** stores static property information and dynamic property information. The static property information is determined when a file is configured and carries over thereafter. Although the static property information can be intentionally altered, it otherwise remains unaltered. The dynamic property information changes over time after a file is created.

[0101] (12) Static Property Information

[0102] The static property information is divided into a file information category and a policy category.

[0103] The file information category includes basic information of a file. In the file information category, a file type indicates the type of the file, such as a text file, document file, picture file, moving picture file or a voice file. An application indicates the application that generated the file. A date created indicates the date the file was first generated. The time at which the file was generated can be registered in addition to the date the file was created. An owner indicates the name of the user who created the file. An access identifier indicates a range of access authorization for the file.

[0104] The policy category is information that is set by the user or the application that created the file, and is information that is designated by the user or the application with regard to file storage conditions. An initial storage class is information that indicates the storage class of the LU in which the file is to be stored when the file is stored in a storage device for the first time. An asset value type indicates the asset value of the file. A life cycle model indicates the model applicable to the file from among life cycle models defined in advance. A migration plan indicates the plan applicable to the file from among plans concerning file migration (hereinafter called "migration") defined in advance.

[0105] The asset value is an attribute that designates the importance or value attached to the file. An attribute of "extra important,""important" or "regular," for example, can be designated as an asset value. The asset value can be used as a supplemental standard for selecting a storage class, i.e., files with an attribute of "important" or higher are stored in LUs that belong to the OnLine Storage class with Premium attribute, or as a standard for selecting a storage class when no life cycle models are designated, for example.

[0106] In the description of the present embodiment, it will be assumed that files that are "important" or higher are stored in LUs that belong to the OnLine Storage (Premium) class. Needless to say, the present invention is not restricted to such an assumption and different standards may be used to select storage classes of LUs for storing files.

[0107] The life cycle stages have been named by drawing analogy with life cycle stages of humans to describe how the usage status of a file changes over time, i.e., the period in which data is created is the birth, the period in which the data is updated and/or used is the growth stage, the period in which the data is rarely updated and is mainly referred to is the mature stage, and the period in which the data is no longer used and is archived is the old age. A life cycle model defines the life cycle a file experiences. The most general method of defining a life cycle is to define the stages based on the amount of time that has elapsed since a file was generated. One example is to define the "growth stage," or the "update stage," in which there are frequent updates, as one month; the "mature stage," or the "reference stage," in which the file is mainly referred to, as one year; and the "old age," or the "archive stage," as thereafter. Hereinafter this definition is called a "model **1**" and is used in the following description. By varying the time interval of the life cycle model or by defining stages with finer resolution, various life cycle models can be defined and one life cycle model from among a plurality of life cycle models can be selected for use. Furthermore, a specific life cycle model can be applied to a certain type of files, or life cycle models can be applied on a per-application basis such that a specific life cycle model is applied to files created by a certain application. Names of the life cycle stages can be expressed in terms of "growth stage,""mature stage," and "old age" that correspond to the life of a person, or in terms of "update stage,""reference stage," and "archive stage" based on file behavior. In the present embodiment, the latter expressions are used in order to more clearly indicate the behavior of files.

[0108] The migration plan defines to which storage class LU a file is transferred according to the file's life cycle stage. One example is a method for storing "update stage" files in OnLine Storage class LUs, "reference stage" files in the NearLine Storage class LUs, and "archive stage" files in Archive Storage class LUs. Hereinafter, this definition is called a "plan **1**" and is used in the following description. In addition to this plan, various plans can be defined, such as a plan that defines "update stage" files to be stored in OnLine Storage (Premium) class LUs, and "reference stage" files in OnLine Storage (Normal) class LUs, while "archive stage" files remain in the NearLine Storage class LUs, and one plan from among a plurality of plans can be selected for use. Furthermore, a specific migration plan can be applied to a certain type of files, or migration plans can be applied on a per-application basis such that a migration plan is applied to files created by a certain application.

[0109] (13) Dynamic Property Information

[0110] The dynamic property information is divided into an access information category and a life cycle information category.

[0111] The access information category includes access statistical information for each file. In the access information category, a time stamp indicates the date and time a given file was last read or written, or the date and time the file storage management table **1100435** of the file was last updated. An access count indicates the total number of accesses to the file. A read count and a write count indicate the number of reads and the number of writes, respectively, to and from the file. A read size and a write size indicate the

average value of the data transfer size when reading and writing, respectively, to and from the file. A read sequential count and a write sequential count indicate the number of times there is address continuity, i.e., sequentiality, between two of multiple consecutive accesses in reading or writing.

[0112] The life cycle information category includes information related to the life cycle of a file. In the life cycle information category, a current life cycle stage indicates the current positioning of a file within its life cycle, i.e., the update stage, the reference stage, or the archive stage. A current storage class indicates the storage class of a storage pool set for the LU that currently stores the file.

[0113] **FIG. 11** indicates one example of the file property information, but various other types of property information can be defined and stored in the file property information management table **1100438**. Furthermore, an embodiment may use only a part of the property information as necessary.

[0114] (14) Initial File Placement: A File Open processing

[0115] Next, a description will be made as to a file open processing that takes place in the initial placement processing to store a file in a storage device for the first time.

[0116] Let us assume that the NAS host **0** (**400**) generated a file abc.doc.

[0117] The NAS host **0** (**400**) issues to the CHN**0** (**1100**) an open request for the file abc.doc. The open request includes a filename as identification information to identify the file. Since the open processing is executed to store the file for the first time, the NAS host **0** (**400**) sends to the CHN**0** (**1100**) the following information included in the file information category and the policy category as the static property information of the file property information, along with the open request. The information sent includes a file type "document," an application that generated the file "XYZ Word," and an access identifier "-rw-rw-rw-" as information included in the file information category, as well as an initial storage class "undesignated," an asset value type "important," the life cycle model "model **1**," and the migration plan "plan **1**" as information included in the policy category.

[0118] The CHN**0** (**1100**) receives the open request from the NAS host **0** (**400**) via the LAN controller **11002**, and the file access control CPU **11001** executes the file system program **110043**.

[0119] When the file system program **110043** is executed, the open request received is specified through a control by the file access control CPU **11001** as an access request to access the local file system LFS**0** (**60**) based on the directory information of the filename. The file open processing section **1100431** refers to the filename management table **1100436** of the LFS**0** (**60**) and searches for abc.doc. Since it is determined as a result that abc.doc is a file that does not yet exist in the filename management table **1100436** and is to be stored for the first time, the file open processing section **1100431** registers abc.doc in the filename management table **1100436** and assigns a file handler to abc.doc.

[0120] Next, the file storage management section **1100433** creates the file storage management table **1100435** to correspond to the file handler assigned to the file abc.doc.

[0121] Next, the file storage management section **1100433** generates the file property information management table **1100438** and correlates it to the file storage management table **1100435** (i.e., a pointer to the file property information management table **1100438** is stored in the file storage management table **1100435**); the file storage management section **1100433** then stores in the file property information management table **1100438** the static property information of the file property information for the file abc.doc obtained from the NAS host **0** (**400**), as well as the date created and owner of the file. Next, the file storage management table **1100435** and the file property information management table **1100438** are written to the LU in which is constructed the file system the file belongs to.

[0122] Next, the CHN**0** (**1100**) returns the file handler to the NAS host **0** (**400**) and the open processing is terminated.

[0123] (15) Initial File Placement: A Data Write Processing

[0124] Next, a description will be made as to a data write processing executed in the initial placement processing of a file.

[0125] Using the file handler obtained in the open processing, the NAS host **0** (**400**) issues to the CHN**0** (**1100**) a write request to store data of the file abc.doc in the storage device **1**.

[0126] When the write request is received by the CHN**0** (**1100**), the file access control CPU **11001** executes the file system program **110043** and uses a method similar to the method used in the open processing to specify that the write request is an access request to access the local file system LFS**0** (**60**).

[0127] The request processing section **1100432** of the file system program **110043** interprets the access request as a write request based on the information included in the access request received, and uses the file handler designated in the write request to obtain the file storage management table **1100435** of the file that corresponds to the file handler.

[0128] Next, the file storage management section **1100433** secures buffers required to store the data and determines the storage positions on disks for the file.

[0129] To determine the storage positions, the file storage management section **100433** refers to the static property information in the file property information management table **1100438**. In this case, due to the fact that the life cycle model of the file abc.doc, which is the subject of the write request, is "model **1**," and to the fact that the write request received is an access taking place within one month of the file generation since it is an access request occurring in an initial file placement, the file storage management section **1100433** specifies the current life cycle stage of the file abc.doc as "growth stage." Further, since the initial storage class is "undesignated" and the asset value type is "important," the file storage management section **1100433** selects "OnLine Storage (Premium)" as the storage class of the storage pool in which to store the file abc.doc.

[0130] Next, the file storage management section **1100433** refers to the storage class management table **1100439** and decides to store the file abc.doc in an LU whose storage class is "OnLine Storage (Premium)" and that is specified by "STR**0** (i.e., the primary storage device **1**)" as the storage node, "FC disk pool **1700**" as the disk pool #, and "LU**0** (i.e., the local file system LFS**0**)" as the LU #. The file storage

management section **1100433** divides the data of the file into one or more logical blocks based on an appropriate algorithm, determines storage addresses of the logical blocks in the LU**0**, generates buffer management tables **1100437** to register the storage addresses determined, and stores in the buffer management table entry of the file storage management table **1100435** pointers to the buffer management tables **1100437** generated. Furthermore, the file storage management section **1100433** stores information in the remaining entries of the file storage management table **1100435**. In the present embodiment, NULL is registered for all entries for link destinations in the file storage management table **1100435**.

[**0131**] The file storage management section **1100433** sets the current life cycle stage as "update stage" and the current storage class as "OnLine Storage (Premium)" in the life cycle information category of the dynamic property information of the file property information management table **1100438**. The file storage management section **1100433** performs appropriate calculations for information included in the access information category of the dynamic property information before registering the results into the file property information management table **1100438**.

[**0132**] The request processing section **1100432** executes a processing according to the write request received; and the LAN controller driver program **110041**, the TCP/IP program **110042**, and the network file system program **110044** are executed by the file access control CPU **11001**; as a result, the write data is transferred from the NAS host **0** (**400**) to the CHN**0** (**1100**) and temporarily stored in the buffer of the file access control memory **11004**. Next, the inter-CPU communications driver program **110046** is executed by the file access control CPU **11001**, and this causes the write request to be transferred to the disk array control CPU **11008** at proper timing. Upon receiving the write request, the disk array control CPU **11008** caches the write data temporarily in the CM **14** and sends a reply of completion with regard to the write request from the NAS host **0** (**400**).

[**0133**] Next, under the control of the DKA **120** that controls disks that make up the LU**0**, the write data is stored at proper timing on appropriate disks.

[**0134**] As described above, files can be initially placed in storage regions that belong to the appropriate storage class based on the static property information of the file.

[**0135**] (16) File Migration Processing (**FIG. 12**)

[**0136**] Next, a migration processing of a file will be described.

[**0137**] The migration management section **110043**A of the file system program **110043** is activated by the file access control CPU **11001** based on a preset timing.

[**0138**] The migration management section **110043**A refers to the file property information management table **1100438** of a file included in the local file system set in advance as the subject of the migration processing, and checks whether the file that is the subject of migration exists. The following is a detailed description of a situation in which the file abc.doc is the subject of the migration processing.

[**0139**] The migration management section **110043**A refers to the file property information management table **1100438** of the file abc.doc and compares the date created to the current date and time. If one month has elapsed since the date created, the migration management section **110043**A recognizes that the current life cycle stage has shifted from the "update stage" to the "reference stage" due to the fact that the life cycle model in the static property information indicates "model **1**" and that one month, which is the period of the "update stage," has already passed.

[**0140**] Further, due to the fact that the migration plan is "plan **1**," the migration management section **110043**A recognizes that the file must be migrated from the LU whose storage class is the "OnLine Storage (Premium)" to an LU whose storage class is the "NearLine Storage."

[**0141**] The migration management section **110043**A refers to the storage class management table **1100439** and decides to transfer the file to an LU whose storage class is the "NearLine Storage" and that is designated by "STR**0** (i.e., the primary storage device **1**)" as the storage node, "SATA disk pool **1710**" as the disk pool #, and "LU**2** (i.e., a local file system LFS**2**)" as the LU #.

[**0142**] Next, the migration management section **110043**A changes the current life cycle stage to "reference stage" and the current storage class to "NearLine Storage" in the dynamic property information of the file property information management table **1100438**.

[**0143**] The migration management section **110043**A defines a unique filename (in this case FILE**00001**) that is used to manage the file abc.doc within the storage device STR**0** (**1**).

[**0144**] The file open processing section **1100431** refers to the filename management table **1100436** of the LFS**2** (**60**) and checks whether the filename FILE**00001** is registered in the filename management table **1100436**; if it is not registered, the file open processing section **1100431** registers the filename FILE**00001** in the filename management table **1100436** and assigns a file handler to the filename FILE**00001**.

[**0145**] Next, the file storage management section **1100433** generates the file storage management table **1100435** and the file property information management table **1100438** to correspond to the file handler assigned to the filename FILE**00001**. Contents identical to the contents registered in the file property information management table of the file abc.doc are stored in the file property information management table **1100438** generated. The file storage management section **1100433** writes in the LU, which stores FILE**00001**, the file storage management table **1100435** and the file property information management table **110438** of FILE**00001**.

[**0146**] Next, the file storage management section **1100433** secures buffer regions required to store the data of FILE**00001** and determines the storage regions (or the storage positions) within the LU**2** for storing the file. Using a method similar to the method used in the data write processing, the file storage management section **1100433** generates the buffer management tables **1100437** to register the storage positions determined, and stores in the buffer management table entry of the file storage management table **1100435** pointers to the buffer management tables **1100437** generated. NULL is registered for all entries for link destinations in the file storage management table **1100435** of the FILE**00001** stored in the LFS**2**.

[0147] As indicated in **FIG. 12**, the file storage management section **1100433** changes the link destination node name to STR**0**, the link destination FS name to LFS**2**, and the link destination filename to FILE**00001** in the file storage management table **1100435** of abc.doc in the LFS**0**.

[0148] Next, the request processing section **1100432** reads data of the abc.doc from disks that make up the LU**0** to buffers in the file access control memory **11004**. The file storage management section **1100433** determines the data read to the buffers in the file access control memory **11004** as data of the FILE**00001** to be written to the disks that make up the LU**2**, and the request processing section **1100432** writes the data to storage regions in the buffers registered in the buffer management tables **1100437**.

[0149] The file storage management section **1100433** clears all buffer management tables **1100437** that can be referred to from pointers registered in the file storage management table **1100435** of the file abc.doc in the LFS**0**, and registers NULL in entries of these buffer management tables **1100437**.

[0150] The data of the FILE**00001** stored in the buffers is stored at proper timing in the LU**2** via the CM **14** of the storage device **1** through a procedure similar to the procedure that took place in the data write processing of the initial placement processing. This completes the migration processing.

[0151] As described above, according to the present embodiment, files can be migrated to storage regions of an appropriate storage class by taking into consideration the life cycle stage of the file based on the migration plan of the file.

[0152] According to the present embodiment, LUs for storing files can be selected based on a concept of storage classes, and LUs for storing files can be changed, without being dependent on host computers or applications executed on the host computers. As a result, a storage device with storage hierarchy, i.e., a plurality of storage regions with varying properties, having high cost effectiveness can be realized without being dependent on host computers.

[0153] Further, due to the fact that data is migrated on a per-file basis, same files can be accessed from a plurality of host computers using a file I/O interface, even after the files are migrated.

[0154] Moreover, a file-based hierarchy storage control can be executed based on static properties of the file, such as the file type, the type of application that generated the file, the intent (policy) of the file generator, and on dynamic properties of the file, such as changes in the life cycle stage, value and access property of the file.

Embodiment 2

[0155] (1) Example of System Configuration (**FIG. 13**)

[0156] Next, referring to **FIG. 13**, an example of the system configuration of the second embodiment will be described. In the present embodiment, a hierarchical storage control is executed between storage devices in a system in which a storage device **1** (hereinafter called "STR**0**") described in the first embodiment and another storage device **1***a* (hereinafter called "STR**1**") are connected via a network.

[0157] In **FIG. 13**, the storage device STR**1** (**1***a*) is the other storage device connected to the storage device STR**0** (**1**) via a LAN **20**; otherwise, the system (configuration components are the same as in **FIG. 1**.

[0158] In the STR**1** (**1***a*), an NCTL**0** (**1100***a*) and an NCTL**1** (**1101***a*) are NAS controllers, and a disk pool **0** (**170***a*) is a disk pool connected to the NCTL**0** and NCTL**1**.

[0159] Instead of the SM I/F control circuit **11005** and the CM I/F control circuit **11006** in the configuration of the CHN **1100** according to the first embodiment shown in **FIG. 4**, the NAS controller NCTLx is provided with an FC controller **11010***a* for connecting with the disk pool **01700***a*. The NAS controller NCTLx also has a cache memory CM **14***a* within the NAS controller, as well as a data transfer control circuit **1011***a*, which is a control circuit for the cache memory CM **14***a*. Further, the data transfer control circuit **11011***a* serves to connect the NAS controller **1100***a* and the NAS controller **1101***a* to each other. Although details of the configuration of the NAS controller NCTL**1** (**1101***a*) are not shown in **FIG. 13**, the NAS controller **1101***a* has a configuration similar to that of the NAS controller **1100***a*. Components that are assigned the same numbers as components of the CHN **1100** in the first embodiment have the same configuration and the same function as the corresponding components of the CHN **1100**.

[0160] Let us assume that the STR**1** is a storage device that is smaller and cheaper than the STR**0**. Also, as shown in **FIG. 13**, a CHN**0** of the STR**0** and the NCTL**0** of the STR**1** are connected via the LAN **20**.

[0161] (2) Migration Processing of File to the Other Storage Device

[0162] The following is a description of the operation according to the present embodiment.

[0163] The CHN**0** (**1100**) of the storage device **1** (STR**0**) recognizes that the storage device **1***a* (STR**1**) of a different type is connected to the LAN **20**. The different storage device can be recognized using a method based on information designated in advance by an administrator or a method based on whether or not there is a device that reacts to a broadcast of a command for recognition to network segments of the LAN **20**. In order to ascertain the configuration of the STR**1**, the CHN**0** of the STR**0** becomes an initiator and issues to the STR**1** a command to collect information. The response from the STR**1** to the command includes the type of the disk pool and the configuration of LUs that the STR**1** has; as a result, by referring to the response, the CHN**0** can recognize that the STR**1** has the SATA disk pool **170***a* and that there is a low-cost file type LU having a 15D +1P configuration RAID **5** and with a large capacity of 2100 GB in the disk pool **170***a*. The CHN**0** of the STR**0** decides to manage the STR**1**'s LU as a remote LU, i.e., as an LU that is in the other storage device STR**1** (**1***a*) but as one of the LUs that are managed by the primary storage device STR**0** (**1**).

[0164] The CHN**0** assigns a number LU**3** to the LU that the STR**1** has and assigns a remote file system number RFS**3** to the file system constructed within the LU. Due to the fact that the LU is in a large capacity, low-cost disk pool, the storage class of the LU is set as "Archive Storage." Based on a control by a disk array control CPU **11008** of the CHN**0**, information regarding the LU**3** in the STR**1**, such as the type of the disk pool, the configuration of the LU, the LU number and the storage class, is stored in a disk pool management

table **131** of an SM **13** of the storage device **1** (STR**0**). The CHN of the storage device **1** refers to the disk pool management table **131** by having a file access control CPU **11001** execute a file system program **110043**, and can register information regarding the LU**3** in a storage class management table **1100451** in a file access control memory **11004** by copying the information regarding the LU**3** from the disk pool management table **131**.

[**0165**] As described in the first embodiment, let us assume that the NAS host **0** (**400**) stored the file abc.doc in the LU**0** of the STR**0** via the CHN**0** and that subsequently the file abc.doc was migrated to the LU**2** of the STR**0** based on a control by the CHN**0**; in the following, only those parts that differ from the first embodiment in the processing executed to migrate the file abc.doc further to the LU**3** in the other storage device STR**1** are described.

[**0166**] As described in the first embodiment, the file abc.doc's current life cycle stage is "reference stage," its current storage class is "NearLine Storage," and its data section is stored under the name FILE**00001** in the LFS**2** of the LU**2** constructed in the SATA disk pool of the STR**0**, as shown in **FIGS. 11 and 12**. The filename management table **1100436** in which the filename "abc.doc" is registered and the file storage management table **1100435** for the file abc.doc are in the LFS**0**. In other words, information regarding the abc.doc is stored in the filename management table **1100436** for the LFS**0** and in the file storage management table **1100435** for the file abc.doc in the LFS**0**. In the meantime, the file property information management table **1100438** is in both the LFS**0** and the LFS**2**. The data section of the file abc.doc has already been migrated to the LU**2** in which is constructed the LFS**2**, which means that the data section of the abc.doc does not reside in the LU**0**, in which is constructed the LFS**0**.

[**0167**] The migration management section **110043**A of the STR**0** refers to the file property information management table **1100438** of the abc.doc and compares the date created to the current date and time. If one year has elapsed since the migration, the migration management section **110043**A recognizes that the current life cycle stage has shifted from the "reference stage" to the "archive stage" due to the fact that the life cycle model in the static property information for abc.doc indicates "model **1**" and that one year, which is the period of the "reference stage," has already passed. Further, due to the fact that the migration plan is "plan **1**," the migration management section **110043**A recognizes that the file must be migrated from an LU whose storage class is "NearLine Storage" to an LU whose storage class is "Archive Storage".

[**0168**] Next, the migration management section **110043**A refers to the storage class management table **1100439**, selects the LU**3** that belongs to the "Archive Storage" class, and decides to transfer the file abc.doc to the LU**3**. The LU**3** has attributes of "STR**1** (i.e., the other storage device la)" as the storage node, "SATA disk pool" as the disk pool #, and "remote file" as the LU type.

[**0169**] Next, the migration management section **110043**A changes the current life cycle stage to "archive stage" and the current storage class to "Archive Storage" in the dynamic property information of the file property information management table **1100438** for the abc.doc.

[**0170**] Next, the migration management section **110043**A defines a unique filename (in this case STR**1**-FILE**00001**) that is used to manage the file abc.doc within the storage device STR**0** (**1**).

[**0171**] The migration management section **110043**A behaves as it were a NAS host and issues to the STR**1** an open request for the file STR**1**-FILE**00001**. This open processing is an open processing executed in order to store the file for the first time from the perspective of the STR**1**. For this reason, the STR**0** includes in the open request sent to the STR**1** the information that the STR**0** has in the file property information management table **1100438** as the static property information of the file abc.doc. However, by changing only the initial storage class in the static property information to "Archive Storage" in the information sent, the STR**0** expressly designates to the STR**1** to store the file STR**1**-FILE**00001** in the Archive Storage class from the beginning.

[**0172**] The NCTL**0** of the STR**1** receives the open request via a LAN controller **11002***a*, and a file access control CPU **11001***a* executes a file system program **110043***a*.

[**0173**] When the file system program **110043***a* is executed, the open request received is specified in a manner similar to the first embodiment as an access request to access the remote file system RFS**3**; the STR**1**-FILE**00001** is registered in a filename management table **1100436***a* in a file access control memory **11004***a* and a file handler is assigned to the STR**1**-FILE**00001** based on a control by the file access control CPU **11001***a*; a file storage management table **1100435***a* and a file property information management table **1100438***a* are created within the file access control memory **11004***a* and information to be registered in the tables is set. The NCTL**0** sends to the migration management section **110043**A of the CHN**0** the file handler assigned to the STR**1**-FILE**00001**, and the open processing is terminated.

[**0174**] Next, like the NAS host **0** (**400**) in the data write processing according to the first embodiment, the migration management section **110043**A of the STR**0** issues to the STR**1** a write request containing the file handler obtained from the NCTL**0** of the STR**1** in the open processing, and requests to write actual data of abc.doc (i.e., data that is also actual data of FILE**00001**) as actual data of the file STR**1**-FILE**00001**.

[**0175**] The file storage management section **1100433***a* of the STR**1** secures buffer regions required to store the write data, determines storage positions on disks of the actual data of the file, and stores the write data received from the STR**0** in the buffers.

[**0176**] To determine the storage positions, the file storage management section **1100433***a* refers to the static property information in the file property information management table **1100438***a*. The file storage management section **1100433***a* specifies the current life cycle stage of the file STR**1**-FILE**00001** as "archive stage" due to the fact that the life cycle model of the file STR**1**-FILE**00001** is "model **1**" and to the fact that more than one year and one month have passed since the file was generated. Further, the file storage management section **1100433***a* specifies "Archive Storage" as the initial storage class as designated by the STR**0**.

[**0177**] The file storage management section **1100433***a* sets the current life cycle stage as "archive stage" and the current storage class as "Archive Storage" in the life cycle infor-

mation category of the dynamic property information of the file property information management table **1100438***a*. The file storage management section **1100433***a* further performs appropriate calculations for access information regarding the file STR**1**-FILE**00001** and updates the information in the access information category of the file property information management table **1100438***a*. NULL is registered for all entries for link destinations in the file storage management table **1100435***a* of the file STR**1**-FILE**00001**.

[**0178**] Next, under the control of the NCTL**0**, the data section of the file STR**1**-FILE**00001** is stored at proper timing on disks that make up the LU**3**.

[**0179**] This concludes the write processing in the STR**1** and the processing returns to STR**0**.

[**0180**] The file storage management section **1100433** of the STR**0** changes the link destination node name to STR**1**, the link destination FS name to LFS**3**, and the link destination filename to STR**1**-FILE**00001** in the file storage management table **1100435** of the FILE**00001** in the LFS**2**. The file storage management section **1100433** then clears all buffer management tables **1100437** that can be referred to from pointers registered in the file storage management table **1100435** of the FILE**00001** and enters NULL in all buffer management table entries of the file storage management table **1100435**.

[**0181**] The preceding transfers the substance of the data section of the file abc.doc from the FILE**00001** in the LFS**2** of the STR**0** to STR**1**-FILE**00001** in the RFS**3** of the STR**1**.

[**0182**] After this, whenever an access request is issued by any of the NAS hosts to access the file abc.doc, the CHN of the STR**0** refers to the file storage management table **1100435** of the abc.doc in the LFS**0** and obtains its link destination node name, FS name and filename, and refers to the file storage management table **1100435** of the FILE**00001** in the LFS**2** based on the identification information of the link destination obtained (i.e., STR**0**, LFS**2**, FILE**00001**). The CHN of the STR**0** further obtains the link destination node name, the FS name and the filename from the file storage management table **1100435** of the FILE**00001** in the LFS**2** and issues to the NCTL of the STR**1** an access request designating identification information of the link destination obtained (i.e., STR**1**, LFS**3**, STR**1**-FILE**00001**), which allows the CHN of the STR**0** to reach STR**1**-FILE**00001** in the RFS**3** of the STR**1** and access the data section of the abc.doc via the NCTL of the STR**1**.

[**0183**] Due to the fact that a plurality of file storage management tables **1100435** must be referred to in order to access a file that has been migrated to the LU**3** of the STR**1** according to the present embodiment, the access speed does suffer slightly. However, since files that are stored in the LU**3** of the STR**1** are files whose current life cycle stage is "archive stage" and therefore files that are rarely subjects of access requests, this poses no problem in practical terms. Even if an access request were to be issued from a host computer for data of a file in "archive stage," the data can be retrieved in real-time from its storage positions on disks since it is stored on magnetic disks, even though it is a file that belongs to the Archive Storage class; unlike conventional situations in which such files are stored on tapes, neither enormous access time for tape control nor transferring the data from the tape to a disk is required according to the present embodiment.

[**0184**] As described above, according to the present embodiment, due to the fact that the storage positions of a file are determined based on the life cycle stage of the file, the Archive Storage class suitable for archiving is selected for files in "archive stage," or the old age, in its life cycle.

[**0185**] Furthermore, other storage devices can be connected to the primary storage device, so that a storage hierarchy that takes advantage of differences in features of various storage devices can be constructed. Files can be migrated to LUs of the other storage devices, instead of migrating only within the primary storage device, according to the migration plan of each file; this further optimizes cost for storage devices compared to situations in which a hierarchical storage control is realized using only one storage device.

[**0186**] In addition, drives on disk devices that make up LUs whose storage class is "Archive Storage" can be halted to realize low power consumption and to extend the life of the disks.

[**0187**] Moreover, due to the fact that even cheaper storage devices can be connected to the storage device STR**1** according to the present embodiment, a storage hierarchy that is even more extensive can be established among a plurality of storage devices; by executing a hierarchical storage control using such a configuration, cost can be further optimized.

Embodiment 3

[**0188**] (1) Example of System Configuration (**FIG. 14**)

[**0189**] Next, with reference to **FIG. 14**, an example of the system configuration according to the third embodiment will be described. In the present embodiment, as in the second embodiment, a hierarchical storage control is executed among storage devices in a system in which another storage device STR**2** (**1***b*) is connected to a storage device STR**0** (**1**) via a network. The third embodiment differs from the second embodiment in that while the network that connects the storage devices was the LAN **20** and file I/O interfaces were used between storage devices in the second embodiment, the network that connects the storage devices is a SAN **35**, which is a dedicated network for connection between the storage devices, and a block I/O interface is used between storage devices in the third embodiment.

[**0190**] In **FIG. 14**, the storage device STR**2** (**1***b*) is a storage device with a small-scale configuration similar to the storage device STR**1** (**1***a*) in the second embodiment, but instead of the NAS controller NCTL**0** of the storage device STR**1** (**1***a*) in the second embodiment, the storage device STR**2** (**1***b*) has SAN controllers FCTLx. Each FCTLx is provided with an FC controller **11012***b* to connect with the SAN **35**, but it does not have the file access control CPU **11001***a* or its peripheral circuits as the STR**1** does and does not perform file control. Otherwise, the storage device STR**2** (**1***b*) according to the present embodiment has a configuration similar to that of the storage device STR**1** (**1***a*) according to the second embodiment.

[**0191**] The SAN **35** is a dedicated network for connecting the storage device STR**0** (**1**) to the storage device STR**2** (**1***b*), and SAN hosts are not connected to the SAN **35**. For the sake of simplification, let us assume that in the present embodiment no SAN hosts are connected to the SAN **35**,

which is the network to connect the storage devices, and that there is only one network that connects the storage devices. However, SAN hosts can be connected to the SAN **35** and a plurality of networks for connecting the storage devices can be provided to improve fault tolerance.

[**0192**] In the present embodiment, the storage device STR**2** (**1***b*) is under the control of the storage device STR**0** (**1**), and file accesses from a NAS host **0** (**400**) reaches the storage device STR**2** (**1***b*) via the storage device STR**0** (**1**). Such a configuration is hereinafter called a "connection of diverse storage devices."

[**0193**] (2) Migration Processing of File to the Other Storage Device

[**0194**] Next, a description will be made as to the processing for migrating a file stored in the STR**0** to the STR**2**, with emphasis on the difference between this processing and the processing according to the second embodiment.

[**0195**] A CHF**1** (**1111**) of the storage device STR**0** (**1**) recognizes that the storage device STR**2** (**1***b*), which is a divergent storage device, is connected to the SAN **35**. The CHF**1** (**1111**) acts as an initiator and issues a command to collect information and thereby recognizes that the STR**2** (**1***b*) is connected to the SAN **35**. The CHF**1** (**1111**) treats storage regions of the STR**2** as if they were a disk pool within the primary storage device according to the first embodiment. A CHN**0** (**1110**) can use the disk pool via the CHF**1** (**1111**). The management method of the disk pool is described later. In order to ascertain the configuration of the STR**2**, the CHN**0** (**1100**) of the STR**0** (**1**) becomes an initiator and issues a command to collect information via the CHF**1** (**1111**) to the STR**2**. The CHN**0** (**1100**) of the STR**0** (**1**) receives a response from the STR**2** to the command via the CHF**1** (**1111**) and recognizes from the information included in the response that the STR**2** has a SATA disk pool and a low-cost, block type LU having a 15D +1P configuration RAID **5** and with a large capacity of 2100 GB; based on this, the CHN**0** (**1100**) of the STR**0** decides to manage the LU as a remote LU. Furthermore, due to the fact that the disk pool that the STR**2** has is a large capacity, low-cost disk pool, the CHN**0** (**1100**) of the STR**0** determines the storage class of the disk pool as "Archive Storage." The CHN**0** (**1100**) of the STR**0** assigns the number LU**4** to the LU inside the STR**2** and stores in a disk pool management table **131** of an SM **13** information concerning the LU, i.e., "Archive Storage" as the storage class #, "STR**2**" as the storage node #, "SATA pool" as the disk pool #, "LU**4**" as the LU #, "remote block" as the LU type, "RAID **5**, 15D +1P" as the RAID Conf., and "2100 GB" as the usable capacity. When the CHN (**1100**) of the STR**0** executes a file system program, the disk pool management table **131** is referred to and the information concerning the LU is copied from the disk pool management table **131** to a storage class management table **1100451** in a file access control memory.

[**0196**] As in the second embodiment, it is assumed that a migration management section **110043**A of the CHN**0** (**1100**) of the STR**0** has decided to migrate a file abc.doc from a NearLine Storage class to the Archive Storage class, and the following is a description of the migration processing of the file executed based on this assumption.

[**0197**] The migration management section **110043**A of the STR**0** refers to a storage class management table **1100439**, selects the LU**4** that belongs to the "Archive Storage" class, and decides to transfer the file abc.doc to the LU**4**. The LU**4** has attributes of "STR**2** (i.e., the other storage device **1***b*)" as its storage node, "SATA disk pool" as its disk pool #, and "remote block" as its LU type.

[**0198**] Unlike the second embodiment, since the LU type of the LU**4** is "block" type, there is no file system in the LU**4**. For this reason, the file system program stored in the CHN**0** (**1100**) constructs a local file system LFS**4** in the LU**4**. Due to the fact that the disk pool in which the LU**4** is set resides in the other storage device STR**2** from the perspective of the STR**0**, it is therefore a "remote" disk pool and the LU**4** is a remote LU; however, since the file system LFS**4** set in the LU**4** is to be controlled by the CHN**0** (**1100**), the file system LFS**4** is managed as a local file system.

[**0199**] Due to the fact that the LFS**4** is to be managed as a local file system and the LU**4** in which the LFS**4** is constructed is an LU that is in the block type, divergent storage device, a file storage management table is treated differently in the present embodiment compared to its treatment in the first and the second embodiments. In other words, a file storage management section **1100433** of the CHN**0** (**1100**) of the STR**0** assigns "STR**2**" as the link destination node name, "LFS**4**" as the link destination FS name, and a STR**2**-FILE**00001** as the link destination filename, and sets these in a file storage management table for the file abc.doc. Since the file abc.doc has already being migrated to the LU**2** under the filename FILE**00001**, the CHN**0** (**1100**) can alternatively set the assigned link destination node name, the link destination FS name and the link destination filename in the file storage management table for the file FILE**00001** in the LFS**2**. Since the STR**2**, in which the LU**4** actually exists, does not execute the file access control as described earlier, a file storage management table for the STR**2**-FILE**00001** is not created in the STR**2**.

[**0200**] The processing that takes place when the file system program **110043** of the CHN**0** (**1100**) is executed is the same as the processing that takes place on the local file system according to the first embodiment in terms of the file open processing, write processing and migration processing, except for the fact that the processing is executed with the awareness that the link destination node of the file abc.doc (i.e., the storage device in which the actual data of the file abc.doc is stored) is the STR**2**.

[**0201**] However, unlike the first embodiment in which a file is transferred to an LU that is in the primary storage device STR**0**, data of a file is transferred to an LU that is in the other storage device STR**2** according to the present embodiment, which results in input/output processing to and from disks that is different from the first embodiment. While the DKA **12***x* of the STR**0** controlled the input/output processing to and from disks in the first embodiment, the CHF**1** (**1111**) of the STR**0** controls the processing according to the configuration of the present embodiment. For this reason, a CHF communications driver program **110096** is stored in a disk array control memory **11009** of the CHN**0**. A CHF communications driver section is realized by having a disk array control CPU **11008** execute the CHF communications driver program **110096**. The CHF communications driver section sends a disk input/output command (hereinafter called an "I/O command") to the SM **13**. Address information representing storage positions of the data is

included in the I/O command. The CHF1 (1111) receives the I/O command via the SM 13 and based on the I/O command received issues an I/O command to the storage device 1*b* (STR2) via the SAN 35. The I/O command issued by the CHF1 (1111) includes address information representing the data storage positions within the storage device 1*b* (STR2). The storage device 1*b* (STR2) processes the I/O command received from the CHF1 (1111) according to the same procedure applied when a disk I/O command is received from a normal host. In other words, the CHF1 of the STR0 is recognized as a host from the perspective of the STR2.

[0202] According to the present embodiment, the disk pool of the divergent storage device STR2 provided with the block I/O interface can be treated as one of the disk pools of the storage device STR0, and a file system managed by the STR0 can be constructed on the LU that is in the disk pool of the STR2. Furthermore, due to the fact that files stored in the LU of the STR0 can be migrated to the LU within the STR2, a flexible storage hierarchy with superior cost effectiveness can be constructed.

Embodiment 4

[0203] (1) Example of System Configuration (FIG. 15)

[0204] The following is a description of the fourth embodiment. The present embodiment differs from preceding embodiments in its configuration.

[0205] FIG. 15 is a diagram of an example of the system configuration according to the present embodiment. A storage device STR3 (1*c*) is provided with a DKC 70 and disk pools. In the DKC 70, an SW 71 is a switch, NNODEs (72*x*) are NAS nodes each provided with a file I/O control mechanism to connect with a LAN, FNODEs (73*x*) are FC nodes each provided with a block I/O control mechanism to connect with a SAN, INODEs (74*x*) are IP nodes each provided with an IP network control mechanism to connect with an IP network, and DNODEs (75*x*) are disk controller nodes each provided with a disk control mechanism to connect with a disk pool. To the switch SW 71 are connected one or more NNODEs 72*x*, one or more FNODEs 73*x*, one or more INODEs 74*x* and one or more DNODEs 75*x*. A node to control iSCSI can be connected to the switch SW 71 to form an IP SAN. The node to control the iSCSI would have functions and a configuration similar to those of the FNODE.

[0206] The DNODE0 and the DNODE1 are connected to and control two types of disk pools, a disk pool 0 and a disk pool 1, which are an FC disk pool 170 and a SATA disk pool 171.

[0207] The INODE2 and the INODE3 are connected to a NAS-type divergent storage device STR1 (1*a*), which is external to the storage device STR3 and is a storage device provided with file I/O interfaces described in the second embodiment. The FNODE2 and the FNODE3 are connected to a SAN-type divergent storage device STR2 (1*b*), which is external to the storage device STR3 and is a storage device provided with block I/O interfaces described in the third embodiment.

[0208] (2) Example of Configuration of the NNODE (FIG. 16)

[0209] FIG. 16 is a diagram of an example of the configuration of the NNODE. The NNODE 720 is equivalent to the CHN 1100 shown in FIG. 4 with the inter-CPU communications circuit 11007 and components below it removed and replaced by an SW node controller 7204. Other components are the same as in the CHN 1100 in terms of configuration and function.

[0210] The SW node controller 7204 is a controller circuit for connecting with the SW 71; it forms commands, data and controller information in internal frame formats that are sent and received within the storage device STR3 (1*c*) and sent as disk I/O to other nodes such as the DNODEs.

[0211] (3) Example of Configuration of the FNODE (FIG. 17)

[0212] FIG. 17 is a diagram of an example of the configuration of the FNODE. The FNODE 730 has a configuration in which an SW node controller 7302 is connected to the FC controller 11012*b* of the FCTL 1100*b* in FIG. 14, which makes the FNODE 730 capable of connecting with the SW 71 via the SW node controller 7302. An FC controller 73-01 operates as a target device and sends and receives frames of commands, data and controller information to and from the SAN. The SW node controller 7302 converts frames sent or received by the FC controller 7301 into internal frame configurations of the storage device STR3 (1*c*) and sends or receives the converted frames to and from other nodes, such as the DNODEs.

[0213] The FNODE 73*x* operates as an initiator device and based on disk I/O commands received from the NNODEs or other FNODEs can send I/O commands to other storage devices connected externally to the storage device STR3. For example, based on commands received from the NNODEs or other FNODEs of the storage device STR3, the FNODE2 and the FNODE3 in FIG. 15 can send I/O commands to the divergent storage device STR2 (1*b*) externally connected to the storage device STR3. In this case, the FNODE2 and the FNODE3 appear to be operating as host computers from the perspective of the STR2.

[0214] Although only the FC controller 7301 and the SW node controller 7302 are shown in FIG. 17 for the sake of simplification, a CPU can be mounted on the FNODEs in order to perform target processing, initiator processing or internal frame generation processing.

[0215] By installing an iSCSI controller instead of the FC controller 7301, a node that controls iSCSI can be configured; by connecting such a node to the SW 71, an IP SAN can be configured.

[0216] (4) Example of Configuration of INODE (FIG. 18)

[0217] FIG. 18 is a diagram of an example of the configuration of the INODE. The INODE 740 has a configuration in which an SW node controller 7402 is connected to a LAN controller 7401, which is similar to the LAN controller 11002*a* of the NCTL0 (1100*a*) in FIG. 13; this configuration makes the INODE 740 capable of connecting with the SW 71 via the SW node controller 7402. The INODEs are provided on the storage device STR3 (1*c*) in order to connect the external NAS-type storage device STR1*a* to the STR3.

[0218] (5) Example of Configuration of DNODE (FIG. 19)

[0219] FIG. 19 is a diagram of an example of the configuration of the DNODE. The DNODE 750 is similar to the

FCTL **1100***b* in **FIG. 14**, but with the FC controller **11012***b* removed and replaced with an SW node controller **7501**. The DNODE **750** goes into operation when it receives a disk I/O command from one of the NNODEs or FNODEs via the SW **71**; as a result, a section **1***d* outlined by a broken line in **FIG. 15** operates as if it were the independent storage device STR**2** in **FIG. 14**. In the present embodiment, the DNODE**0** (**750**) and the DNODE**1** (**751**) form a pair of redundant controllers. Having redundant DNODEs is similar to the configuration of the storage device STR**2** in **FIG. 14**, where there are also redundant FCTLs.

[**0220**] (6) Migration Processing of Files

[**0221**] The present embodiment only differs from the first, second and third embodiments in its configuration of the storage device, and its processing procedure for executing a hierarchical storage control is similar to that in the first, second and third embodiments; accordingly, only those parts that differ in the operation as a result of differences in the configuration of the storage device are described below.

[**0222**] In the present embodiment, a hierarchical storage control inside the storage device STR**3** can be executed using a procedure similar to that in the first embodiment. A file system program **110043** stored in a file access control memory of the NNODE **72***x* is equipped with a storage class management table **1100439** for managing usable LUs, and can recognize disk pools and LUs managed by the DNODEs **75***x* by referring to the storage class management table **1100439**. However, unlike the first embodiment, there is no SM **13** for storing shared information; consequently, the NNODE **72***x* must query all DNODEs **75***x* in advance to specify a usable LU and register it in the storage class management table **1100439**. Of course, an SM node for connecting with an SM can be provided for connection with the SW **71** in the present embodiment, so that the storage class management table **1100439** can consist of information stored in the SM, as in the first embodiment.

[**0223**] When the NNODE **72***x* specifies a usable disk pool and an LU, creates the storage class management table **1100439**, and defines a storage class, a processing similar to that in the first embodiment can be applied subsequently to execute a hierarchical storage control within the storage device STR**3** (**1***c*), i.e., a hierarchical storage control using LUs set in the disk pool **0** and the disk pool **1**.

[**0224**] To issue a disk I/O command, an SW node driver program stored in the file access control memory **7203** of the NNODE is executed by the file access control CPU **7202**, which causes a disk I/O command to be issued via the SW node to the DNODE **750** that manages the LU that is the subject of access.

[**0225**] Through the configuration and processing described above, a system in which a file-based storage hierarchy is constructed within the storage device STR**3**, as in the first embodiment, can be realized.

[**0226**] Furthermore, the NAS-type divergent storage device STR**1** (**1***a*) provided with file I/O interfaces can be connected externally to the storage device STR**3**, which would result in a configuration of a storage hierarchy as in the second embodiment. When the file system program **110043** stored in the file access control memory **7203** of the NNODE **72***x* is executed by the file access control CPU **7202**, the file system program **110043** queries the INODE **74***x* whether there is a NAS-type divergent storage device connected to the INODE **74***x*; if there is a divergent storage device connected, the file system program **110043** obtains from the divergent storage device the information for identifying remote LUs and remote file systems that are in the divergent storage device. Through a control by the file access control CPU **7202**, a storage class is defined for each of the remote LUs and the remote file systems, and information concerning the LUs is registered and managed in the storage class management table **1100439**. Subsequent steps are the same as in the processing procedure in the second embodiment.

[**0227**] To issue a disk I/O command, the SW node driver program stored in the file access control memory **7203** of the NNODE is executed by the file access control CPU **7202**, which causes a disk I/O command to be issued from the NNODE via the SW node to the INODE **740** connected to the storage device STR**1** (**1***a*) that is provided with the LU that is the subject of access. Based on the I/O command received, the INODE **740** issues to the storage device STR**1** (**1***a*) a disk I/O command for a file access, as well as sends and receives actual data of the file and control information to and from the STR**1** (**1***a*).

[**0228**] The INODEs **74***x* have no involvement whatsoever in the file control information and operate simply as gateways of an IP network. In such a case, a hierarchical storage configuration without any interference from other devices, such as NAS hosts, can be realized. Of course, the divergent storage device STR**1** (**1***a*) can be connected to the LAN **20**, to which the NNODE **720** is connected, as in the second embodiment.

[**0229**] Through the configuration and processing described above, a file-based storage hierarchy that utilizes storage pools of an external divergent storage device, as in the second embodiment, can be realized.

[**0230**] Furthermore, the SAN-type divergent storage device STR**2** (**1***b*), which is a storage device provided with block I/O interfaces, could be connected externally to the storage device STR**3**, which would result in a configuration of a storage hierarchy as in the third embodiment. When the file system program **110043** stored in the file access control memory **7203** of the NNODE **72***x* is executed by the file access control CPU **7202**, the NNODE queries the FNODEs **73***x* whether there is a SAN-type divergent storage device connected to the FNODEs **73***x*. If there is a divergent storage device connected, the NNODE recognizes remote LUs of the divergent storage device based on the contents of the response from the FNODEs **73***x* to the query and constructs local file systems in the remote LUs. The NNODE then defines a storage class for each of the remote LUs and the local file systems, and registers and manages information concerning the LUs in the storage class management table **1100439**. Subsequent steps are the same as in the third embodiment.

[**0231**] To issue a disk I/O command, the SW node driver program is executed by the file access control CPU **7202**, which causes a disk I/O command to be issued from the NNODE via the SW node to the FNODE **732** connected to the storage device STR**2** (**1***b*), which is provided with the LU that is the subject of access. The FNODE **732** issues to the storage device STR**2** a disk I/O command, as well as sends and receives data and control information to and from the

STR2. Through the configuration and processing procedure described above, a file-based storage hierarchy that utilizes a file system that is constructed in an external storage device STR2 as in the third embodiment and managed by the STR3 can be realized.

[0232] According to the present embodiment, the storage device STR3 behaves as if it were a central control controller for constructing a hierarchy storage system and various types of storage devices can be connected internally and externally to the storage device STR3; consequently, an extremely flexible, scalable and large-scale hierarchical storage system can be constructed. Furthermore, due to the fact that disks and other storage devices can be connected internally and externally to the storage device STR3 as nodes on the SW 71 of the storage device STR3, high-speed data transfer becomes possible.

[0233] (7) Other Applications

[0234] Although file transfer methods and storage devices that execute hierarchical migration processing of files based on file's data life cycle stages have been described in the first through fourth embodiments, files can be transferred based on other standards and a plurality of standards can be combined. Possible standards other than the data life cycle stage include a file's access property and an LU's used capacity. In such cases, transfer of files can be controlled by providing a migration plan based on the file's access property or the LU's used capacity.

[0235] Examples of migration plans based on a file's access property include a plan to re-transfer a file into a storage class one class higher in the hierarchy when the access frequency of the file exceeds a certain level, or a plan that provides a storage class specialized for sequential accesses and that transfers a file into the specialized storage class once the sequential access frequency to the file exceeds a certain level.

[0236] Examples of migration plans based on an LU's used capacity include a plan to transfer a file to a storage class one class lower in the hierarchy even if its current life cycle stage has not shifted, if the file that is stored in an LU has low access frequency or if a long time has elapsed since its date of creation, once the used capacity of the LU exceeds a certain level.

[0237] The file property information management table 1100438 in the above embodiments manages dynamic properties for each file as access information. The storage class management table 1100439 manages the total capacity and used capacity of each LU. By utilizing such information, migration plans described above can be readily realized.

[0238] A hierarchical storage control according to the property of a file can be realized through a processing within a storage device and without being dependent on host computers.

[0239] While the description above refers to particular embodiments of the present invention, it will be understood that many modifications may be made without departing from the spirit thereof. The accompanying claims are intended to cover such modifications as would fall within the true scope and spirit of the present invention.

[0240] The presently disclosed embodiments are therefore to be considered in all respects as illustrative and not restrictive, the scope of the invention being indicated by the appended claims, rather than the foregoing description, and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced therein.

What is claimed is:

1. A storage system that is connected to at least one computer, the storage system comprising:

a first interface control device that receives from the at least one computer an access request designating identification information of a file;

a second interface control device connecting to the first interface control device; and

a plurality of disks connecting to the second interface control device, wherein the plurality of disks include at least one first disk, and at least one second disk, the first disk and the second disk being different kinds, the first interface control device decides based on identification information received from the computer a storage position of data of the file designated by the identification information within the plurality of disks, and

the second interface control device controls to store the data of the file designated by the identification information at the storage position decided by the first interface control device.

2. A storage system according to claim 1, wherein the at least one first disk is a Fibre Channel disk equipped with a Fibre Channel type interface, and the at least one disk is a serial ATA disk equipped with a serial ATA type interface.

3. A storage system according to claim 1, further comprising a memory, a memory controller for controlling the memory, a plurality of first interface control devices each being connected to the memory controller, and a plurality of second interface control devices each being connected to the memory controller,

wherein one of the first interface control devices receives identification information of a file and data of the file from the computer, and stores the data of the file in the memory, and

one of the second interface control devices that is connected to one of the disks that is to store the data of the file controls to store the data of the file retained in the memory in the one of the disks according to the storage position of the data of the file decided by the first interface control device.

4. A storage system according to claim 1, wherein

a first storage region exists in the at least one first disk,

a second storage region exists in the at least one second disk, and

the first interface control device sets up a first file system in the first storage region, and a second file system in the second storage region.

5. A storage system according to claim 4, wherein the first interface control device decides, according to static property that is predetermined property and dynamic property that is property that changes with passage of time since a point of time when the file is created, as to which one of the first storage region and the second storage region to store the data of the file indicated by the identification information.

6. A storage system according to claim 5, wherein the first interface control device controls to migrate data of a file stored in one of the first storage region and the second storage region to the other storage region according to a change in the dynamic property, and changes identification information that specifies the file and information indicative of correlation between the file and the storage position.

7. A storage system according to claim 6, wherein the static property includes at least one of information that specifies a kind of the file, information that specifies a time when the file is created, and information that specifies a value of the file, and the dynamic property includes at least one of information concerning an access property to the file, and information concerning passage of time elapsed since the file is created.

8. A storage system that is connected to a computer, the storage system comprising:

at least one first interface control device that connects to the computer and receives from the computer an access request containing identification information of a file;

at least one second interface control device that connects to the at least one first interface control device; and

a plurality of first disks each being connected to the at least one second interface control device,

wherein the at least one first interface control device connects to a second storage system having a plurality of second disks,

a first storage region is set in the plurality of first disks,

a second storage region is set in the plurality of second disks,

the at least one first interface control device, upon receiving an access request from the computer, decides, according to property of a file designated by identification information contained in the access request received, as to which one of the first storage region and the second storage region to store data of the file,

when the data of the file is stored in the first storage region, the at least one second interface control device stores the data of the file in one of the plurality of first disks, and

when the data of the file is stored in the second storage region, the at least one first interface control device that received the access request from the computer controls such that the data of the file is transmitted to the second storage system through the at least one first interface control device that is connected to the second storage system.

9. A storage system according to claim 8, wherein

the second storage system includes a third interface control device that receives an access request having identification information of a file, and that accesses a storage region within the second storage region correlated to the identification information received to thereby access data of the file specified by the identification information.

10. A storage system according to claim 9, wherein the third interface control device sets a file system in the second storage region, and wherein, when the first interface control device receives an access request from the computer, and when data of a file designated by identification information contained in the access request is to be stored in the second storage region, the first interface control device controls to transmit the access request having the identification information correlated to the file to the third interface control device through the first interface control device connected to the second storage system.

11. A storage system according to claim 10, wherein the first interface control device connected to the second storage system receives an access request from the computer.

12. A storage system according to claim 10, wherein the first interface control device migrates the data of the file from the first storage region to the second storage region through the third interface control device, based on property of files whose data is stored in the first storage region.

13. A storage system according to claim 12, wherein, when the first interface control device migrates data of a file stored in the first storage region to the second storage region, the first interface control device transmits to the third interface control device an access request having identification information correlated to the file, and stores the identification information of the file received from the computer correlated with the file system set in the second storage region.

14. A storage system according to claim 13, wherein

the first interface control device stores information concerning property of files and information concerning property of the first storage region and the second storage region, and decides, based on the information concerning property of files and the information concerning property of the first storage region and the second storage region, as to whether or not data of a file stored in the first storage region is to be migrated to the second storage region.

15. A storage system that is connected to a computer, the storage system comprising:

a first interface control device that receives from the computer an access request having identification information for designating a file;

a second interface control device that is connected to the first interface control device;

a plurality of first disks that are connected to the second interface control device;

a third interface control device that is connected to a second storage system having a fourth interface control device that receives an access request containing address information indicating a storage position of data and having a plurality of second disks that are connected to the fourth interface control device;

a first storage region existing in the plurality of first disks; and

a second storage region existing in the plurality of second disks; wherein

the first interface control device, upon receiving an access request for a file from the computer, decides as to which one of the first storage region and the second storage region to store data of the file according to property of the file indicated by identification information contained in the access request received,

when the data of the file is stored in the first storage region, the second interface control device stores the data of the file in one of the plurality of first disks, and

when the data of the file is stored in the second storage region, the first interface control device controls to transmit to the fourth interface control device through the third interface control device the access request containing address information for an address within the second storage region where the data of the file is to be stored.

16. A storage system according to claim 15, wherein the third interface control device is an interface control device that corresponds to a block I/O interface.

17. A storage system according to claim 16, wherein the first interface control device sets up a file system in the second storage region.

18. A storage system according to claim 17, wherein the first interface control device controls to migrate data of a file from the first storage region to the second storage region through the third interface control device based on property of the file whose data is stored in the first storage region.

19. A storage system according to claim 18, wherein, when the first interface control device migrates the data of the file stored in the first storage region to the second storage region, the first interface control device controls to transmit to the second storage system through the third interface control device an access request containing an address of a storage region in the second storage region that stores the data of the file, and changes relation between identification information of the file and information indicating the storage region that stores the data of the file.

20. A storage system that is connected to a computer, the storage system comprising:

a first node that receives from the computer an access request containing identification information of a file;

a second node that is connected to at least one first disk;

a third node that is connected to at least one second disk and a second storage system that is connected to the at least one second disk and has a file I/O interface control device that receives an access request having identification information of a file;

a fourth node that is connected to at least one third disk and a third storage system that is connected to the at least one third disk and has a block I/O interface control device that receives an access request having address information of the file indicating a storage position of the file within the at least one third disk; and

a switch device that mutually the first node, the second node, the third node and the fourth node,

wherein a first storage region exists in the at least one first disk, a second storage region exists in the at least one second disk, and a third storage region exists in the at least one third disk, and

the first node controls to store data of the file in one of the first storage region, the second storage region and the third storage region according to property of the file specified by identification information received from the computer.

21. A storage system according to claim 20, wherein, when the data of the file is stored in the first storage region, the second node controls to store the data of the file in a storage region in the at least one first disk.

22. A storage system according to claim 20, wherein, when the data of the file is stored in the second storage region, the first node controls to transmit to the second storage system through the third node an access request containing identification information correlated to the file.

23. A storage system according to claim 20, wherein, when the data of the file is stored in the third storage region, the first node controls to transmit to the third storage system through the fourth node an access request containing address information indicating a storage position of the data of the file.

24. A storage system according to claim 20, wherein,

when the data of the file is stored in the first storage region, the second node controls to store the data of the file in a storage region in the at least one first disk,

when the data of the file is stored in the second storage region, the first node controls to transmit to the second storage system through the third node an access request containing identification information correlated to the file, and

when the data of the file is stored in the third storage region, the first node controls to transmit to the third storage system through the fourth node an access request containing address information indicating a storage position of the data of the file.

25. A storage system comprising:

a file I/O interface control device; and

a plurality of disk pools,

wherein the file I/O interface control device sets one of a plurality of storage hierarchies defining storage classes, respectively, for each of LUs within the disk pools, thereby forming a file system in each of the LUs, and the file I/O interface control device migrates at least one of the files from one of the LUs to another one of the LUs of an optimal storage class, based on static properties and dynamic properties of each file.

* * * * *