

(19) 世界知的所有権機関
国際事務局(43) 国際公開日
2008 年 7 月 31 日 (31.07.2008)

PCT

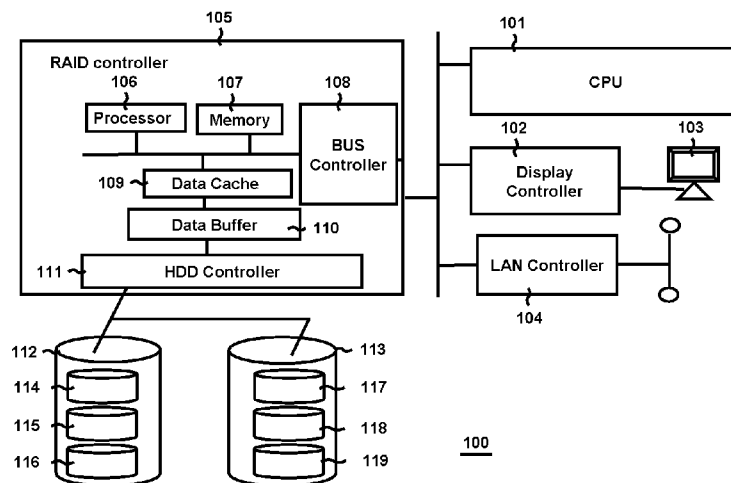
(10) 国際公開番号
WO 2008/090620 A1

- (51) 国際特許分類:
G06F 3/06 (2006.01)
- (21) 国際出願番号: PCT/JP2007/051195
- (22) 国際出願日: 2007 年 1 月 25 日 (25.01.2007)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人 (米国を除く全ての指定国について): 富士通株式会社 (FUJITSU LIMITED) [JP/JP]; 〒2118588 神奈川県川崎市中原区上小田中 4 丁目 1 番 1 号 Kanagawa (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 櫻井 博 (SAKURAI, Hiroshi) [JP/JP]; 〒2118588 神奈川県川崎市中原区上小田中 4 丁目 1 番 1 号 富士通株式会社内 Kanagawa (JP).
- (74) 代理人: 横山 淳一 (YOKOYAMA, Junichi); 〒2118588 神奈川県川崎市中原区上小田中 4 丁目 1 番 1 号 富士通株式会社内 Kanagawa (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY,

[続葉有]

(54) Title: STORAGE DEVICE, RECORD RECOVERY METHOD, AND RECORD RECOVERY PROGRAM

(54) 発明の名称: 記憶装置、記録復旧方法、記録復旧プログラム



(57) Abstract: [PROBLEMS] To obtain data recorded on disks in which failures have occurred without fail by inhibiting the writing of data involved in startup even if, in a storage device comprising plural disks, failures have occurred thereon. [MEANS FOR SOLVING PROBLEMS] A storage device is characterized by having plural recording media and comprising a judging means for judging whether or not the number of stopped recording media is more than one in the storage device allowing the stop of the recording media in which failures have been detected, a startup means for restarting the storage device having judged the stop of the recording media, a storage means for storing initial data involved in the startup of the storage device, an operating means for operating any one of the stopped recording media, a specifying means for inhibiting the writing of the initial data on the operated recording medium and specifying the operation order of the storage device to be operated, and a writing means for recording the initial data stored in the storage means on a recording medium which has been judged that the operation order is correct in the judging means.

(57) 要約: (課題) 本発明は、複数のディスクから構成される記憶装置において、複数のディスクに障害が発生した場合であっても、起動に伴うデータの書き込みを禁止し、障害の発生したディスクに記録されたデータを確実に取得することを目的とす

[続葉有]

WO 2008/090620 A1



KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:
— 国際調査報告書

る。(解決手段) 本実施例に係る記憶装置は、複数の記録媒体を有し、障害を検出した記録媒体を休止する記憶装置において、休止した記録媒体が複数であるか否かを判別する判別手段と、複数の記録媒体の休止を判別した該記憶装置を再起動する起動手段と、該記憶装置の起動に伴う初期データを格納する格納手段と、休止した記録媒体のいずれか一つを稼動する稼動手段と、該稼動した記録媒体に該初期データの書込みを禁止し、稼動する記憶装置の稼動順序を特定する特定手段と、該判別手段において、稼動順序が正しいと判別した記録媒体に対して、該格納手段に格納した該初期データを記録する書込み手段とからなることを特徴とする。

明 細 書

記憶装置、記録復旧方法、記録復旧プログラム

技術分野

[0001] 冗長構成の記憶装置において、記録されたデータの復旧を行う記憶装置に関する。

背景技術

[0002] 複数のディスクから構成されるRAID装置などの冗長化された記憶装置は、ディスクに障害が発生すると、障害の発生したディスクを切り離す。ここでディスクの障害としては例えば、サーマル・オフトラックやコンタミネーション、ノイズ、接触不良などがある。

[0003] 記憶装置は、障害の発生したディスクを切り離し、正常なディスクのみを用いてデータの記録更新を行う。さらに他のディスクに障害が発生すると記憶装置は、新たに障害の発生したディスクを切り離す。複数のディスクに障害が発生し、冗長が無くなった状態から更にディスクを切り離した状態を以下、マルチデッド状態と呼ぶ。ここで、障害発生して切り離されたディスクは、ハードリセットや電源の再投入を行と一時的に正常なディスクとして動作することができる。これはディスクの切り離し要因として、一時的なノイズや、サーマル・オフトラックやディスク内媒体上の微小ゴミ(コンタミネーション)などの場合が多く、電源切断後の投入や、ハードリセットなどによって回復することが多い。このため、マルチデッド状態から冗長化を失った直前の状態に記憶装置を復旧するために、一度切り離されたディスクに対してハードリセットや電源の再投入を行う。そして、切り離されたディスクを稼動して冗長化を失った直前の状態にする。しかし、記憶装置を再起動する場合、複数のディスクに対して冗長化を失った直前の状態を確実に判断できない場合、ディスクへの誤書込みが発生し、システムが起動しなくなったり、誤動作する可能性があった。つまり、一度切り離された複数のディスクを復旧するために、再接続すべきディスクの順番を特定できない場合、冗長化を失った直前の状態を確実に判断できないといった問題がある。

[0004] またディスク装置の復旧可能なディスクアレイ制御装置に関して以下の特許文献が

ある。

特許文献1:特開平10-289065号公報

発明の開示

[0005] (発明が解決しようとする課題)

複数のディスクから構成される記憶装置において、複数のディスクに障害が発生し、冗長化を失った直前の状態を確実に判断できない場合でも、ディスクへの誤書込みをせず、冗長化を失った直前の状態まで該記憶装置を復旧し、正常に起動することを目的とする。

(課題を解決するための手段)

本発明の一観点によれば、本発明の記憶装置は、複数の記録媒体を有し、障害を検出した記録媒体を休止する記憶装置において、休止した記録媒体が複数であるか否かを判別する判別手段と、複数の記録媒体の休止を判別した該記憶装置を再起動する起動手段と、該記憶装置の起動に伴う初期データを格納する格納手段と、休止した記録媒体のいずれか一つを稼動する稼動手段と、該稼動した記録媒体に該初期データの書込みを禁止し、稼動する記憶装置の稼動順序を特定する特定手段と、該特定手段において、稼動順序が正しいと判別した記録媒体に対して、該格納手段に格納した該初期データを記録する書込み手段とからなることを特徴とする。

[0006] また本発明の他の観点によれば、本発明の記憶装置は、該特定手段が稼動した記録媒体に記録される履歴情報を参照して該稼動順序を特定することを特徴とする。

[0007] また本発明の他の観点によれば、本発明の記憶装置は、該特定手段が該記憶装置の動作設定を参照して該稼動順序を特定することを特徴とする。

[0008] また本発明の他の観点によれば、本発明の記憶装置は、該特定手段が該記憶装置が有するオペレーションシステムの動作状態を参照して該稼動順序を特定することを特徴とする。

[0009] また本発明の他の観点によれば、本発明の記憶装置は、該複数の記録媒体に同一のデータを格納することを特徴とする。

[0010] また本発明の他の観点によれば、本発明の記憶装置は、該複数の記録媒体にデータを分散して記録し、さらに分散した分散データの排他的論理和を格納することを

特徴とする。

[0011] また本発明の他の観点によれば、本発明の記録復旧方法は、複数の記録媒体を有し、障害を検出した記録媒体を休止する記憶装置が記録するデータを復旧する記録復旧方法において、休止した記録媒体が複数であるか否かを判別する判別手順と、複数の記録媒体の休止を判別した該記憶装置を再起動する起動手段と、該記憶装置の起動に伴う初期データを格納する格納手順と、休止した記録媒体のいずれか一つを稼動する稼動手順と、該稼動した記録媒体に該初期データの書込みを禁止し、稼動する記憶装置の稼動順序を特定する特定手順と、稼動順序が正しいと判別した記録媒体に対して、該格納手段に格納した該初期データを記録する書込み手順とからなることを特徴とする。

[0012] また本発明の他の観点によれば、本発明の記録復旧プログラムは、複数の記録媒体を有し、障害を検出した記録媒体を休止する記憶装置が記録するデータを復旧する記録復旧プログラムにおいて、休止した記録媒体が複数であるか否かを判別する判別ステップと、複数の記録媒体の休止を判別した該記憶装置を再起動する起動ステップと、該記憶装置の起動に伴う初期データを格納する格納ステップと、休止した記録媒体のいずれか一つを稼動する稼動ステップと、該稼動した記録媒体に該初期データの書込みを禁止し、稼動する記憶装置の稼動順序を特定する特定ステップと、稼動順序が正しいと判別した記録媒体に対して、該格納手段に格納した該初期データを記録する書込みステップとをコンピュータに実行させることを特徴とする。

(発明の効果)

本実施例によれば、複数のディスクから構成される記憶装置において、再起動時の書き込みデータをバッファに保存して、障害の発生したハードディスクへの書込みを禁止することによって、複数のディスクに障害が発生した場合であっても、障害の発生したハードディスクの順序を特定し、該記憶装置を正常に起動することができる。

図面の簡単な説明

[0013] [図1]本実施例に係るRAID装置100のハードブロック図である。

[図2]本実施例に係る記録処理のフローチャートである。

[図3]本実施例に係る記録復旧のフローチャートである。

[図4]本実施例に係るシステム領域114、117に格納されるデータ401、402である。

[図5]本実施例に係るシステム領域114、117に格納されるデータ501、502である。

[図6]本実施例に係るRAID管理領域116、119に格納されるデータ601、602である。

[図7]本実施例に係るRAID管理領域116、119に格納されるデータ701、702である。

[図8]本実施例に係るRAID管理領域116、119に格納されるデータ801、802である。

[図9]本実施例に係るRAID装置900のハードブロック図である。

[図10]本実施例に係る記録処理のフローチャートである。

[図11]本実施例に係る記録復旧のフローチャートである。

[図12]本実施例に係るシステム領域915、918に格納されるストライプデータ1201、1202である。

[図13]本実施例に係るシステム領域921に格納されるパリティデータ1301である。

[図14]本実施例に係るRAID管理領域917、920に格納される管理データ1401、1402、1407である。

[図15]本実施例に係るRAID管理領域917、920に格納される管理データ1601、1602、1607である。

[図16]本実施例に係るRAID管理領域917、920に格納される管理データ1701、1702、1707である。

[図17]本実施例に係るメモリ107に格納される設定テーブル1801である。

[図18]本実施例に係るメモリ107に格納される設定テーブル1901である。

[図19]本実施例に係るメモリ907に格納される設定テーブル2001である。

符号の説明

- [0014] 100…RAID装置
- 101…CPUモジュール
- 102…ディスプレイコントローラ
- 103…ディスプレイ

104…LANコントローラ
105…RAIDコントローラ
106…プロセッサ
107…メモリ
108…バスコントローラ
109…データキャッシュ
110…データバッファ
111…HDDコントローラ
112…ハードディスク
113…ハードディスク
114…システム領域
115…データ領域
116…RAID管理領域
117…システム領域
118…データ領域
119…RAID領域
401…データ
402…データ
403…設定テーブル
404…設定テーブル
601…データ
602…データ
603…管理情報
604…管理情報

発明を実施するための最良の形態

[0015] (実施例1)

本実施例は記憶装置100がミラーリング構成の場合を説明する。ミラーリング構成とは、2以上の記録媒体が共に同一のデータを記録する構成である。

[0016] [システム構成]

図1は本実施例に係る記憶装置100のハードブロック図である。記憶装置100は例えば、RAID (Redundant Arrays of Inexpensive Disks) 装置などの記憶装置である。本実施例では、以下記憶装置100をRAID装置100と呼ぶ。

[0017] <RAID装置100>

RAID装置100は、CPU (Central Processing Unit) モジュール101、ディスプレイコントローラ (Display Controller) 102、ディスプレイ103、LANコントローラ (LAN Controller) 104、RAIDコントローラ (RAID Controller) 105、ハードディスク (HDD) 112、113から構成される。CPUモジュール101は、RAID装置100を統括的に制御し、ディスプレイコントローラ102、LANコントローラ104、RAIDコントローラ105を統括的に制御する。またCPUモジュール101はメモリなどの記録媒体を含む構成であり、RAID装置100のOSをメモリ上に展開してOSを起動するものである。ディスプレイコントローラ102は、ディスプレイ103における表示制御を行う。LANコントローラ104はRAID装置100が接続するネットワークの通信制御を行う。RAID装置100は、LANコントローラ104を介してRAID装置100の外部のネットワークにより、所望のデータを取得する。

[0018] <RAIDコントローラ105>

RAIDコントローラ105は、プロセッサ106、メモリ107、バスコントローラ108、データキャッシュ109、データバッファ110、HDDコントローラ111から構成されている。プロセッサ106は、HDDコントローラ111経由でハードディスク112またはハードディスク113にデータを書き込む／読み出す制御を行う。プロセッサ106はメモリ107に格納されているプログラムを実行する。バスコントローラ108は、RAIDコントローラ105とCPUモジュール101でのデータのやりとりを制御する。さらにプロセッサ106はHDDコントローラ111経由でハードディスク112、113との接続制御を行う。RAIDコントローラ105とハードディスク112、113を接続することを「オンライン」、RAIDコントローラ105とハードディスク112、113を切断することを「オフライン」と呼ぶ。ただしオフライン、オンラインに限らず、プロセッサ106はHDDコントローラ111経由でハードディスク112、113から管理情報を読み出すことが可能である。データキャッシュ109は

ハードディスク112、113のアクセスを補助し、一度ハードディスク112、113にアクセスするデータを一時的に保管し、保管された情報は、ディスクに直接アクセスせずに高速化する。またデータバッファ111は、ハードディスク112、113に書き込むデータを一旦、格納する記憶部である。ハードディスク112、113は一定時間に一定のデータ量を書き込む必要があるために、データバッファ110は設けられている。換言すれば、データキャッシュ109とデータバッファ110は、CPUモジュール101またはプロセッサ106が書き込みを指示するデータをハードディスク112、113に格納する前に格納する。そしてデータ量が所定の容量よりも大きくなった後、HDDコントローラ111は、データキャッシュ109経由でデータバッファ110からデータを読み出し、ハードディスク112、113に格納する。データキャッシュ109とデータバッファ110があることによって、プロセッサ106はデータの転送をハンドリングすることができる。

[0019] ハードディスク112はシステム領域114、データ領域115、RAID管理領域116より構成されており、ハードディスク113はシステム領域117、データ領域118、RAID管理領域119より構成されている。システム領域114、データ領域115、RAID管理領域116は、ハードディスク112を分割した領域である。同様にしてシステム領域117、データ領域118、RAID管理領域119は、ハードディスク113を分割した領域である。そしてシステム領域114、117、データ領域115、118、RAID管理領域116、119の容量は可変な量であって、ハードディスク112、113のデータ容量の範囲内で可変である。

[0020] システム領域114、117にはRAID装置100におけるOS (Operation System)、アプリケーションプログラムなどが格納されている。データ領域115、118にはRAID装置100のデータが格納されている。RAID管理領域116、119にはRAID装置100の管理情報が格納されている。ここでデータとはユーザの個人情報などであって、RAID装置100を制御するプログラムなどとは異なる情報である。管理情報とは、ハードディスク112、113が正常か否かを示す情報であり、プロセッサ106がHDDコントローラ111経由で更新を行う。

図4(a)は、本実施例に係るシステム領域114に格納されているデータ401である。データ401は、RAID装置100において起動するOS404、アプリケーションプログラ

ム406から構成されている。同様にして図4(b)は、本実施例に係るシステム領域117に格納されているデータである。データ402は、RAID装置100において起動するOS405、アプリケーションプログラム407から構成されている。

また図17に記載の設定テーブル1801はハードディスク112、113、データキャッシュ109、データバッファ110へのデータの書き込み／読み出しを決定する書き込みフラグのオン／オフを示すテーブルである。初期設定では設定テーブル403の書き込みフラグはすべてオンになっている。設定テーブル1801は、メモリ107に格納されている。

本実施例において、RAID装置100はミラーリング構成であるため、データ401を構成するOS404とデータ402を構成するOS405、データ401を構成するアプリケーションプログラム406とデータ402を構成するアプリケーションプログラム407は同一のOS及びアプリケーションプログラムである。

本実施例においてBIOS (Basic Input/Output System) 設定などにおいて、ユーザはデータキャッシュ109、データバッファ110、ハードディスク112、113の書き込みを有効/無効にするフラグのオン／オフを設定する。

またハードディスク112、113への書き込みを無効にした場合、プロセッサ106はデータをデータキャッシュ109またはデータバッファ110から直接読み出しが可能とする。プロセッサ106は設定テーブル1801を図18に記載の設定テーブル1901に更新する。ハードディスク112、113へデータの書き込みフラグをオフにする。そしてデータキャッシュ109へデータの書き込みを許可する書き込みフラグをオンのままにして変更せず(1にする。書き込みフラグを立てる。)、データバッファ110への書き込みを許可する書き込みフラグもオンのままにする。

図6(a)は、RAID管理領域116に格納されているデータ601である。データ601は、管理情報603とシステムの起動時間から構成されている。管理情報603は、ハードディスク112、113に障害が発生していない場合の管理情報であり、ハードディスク112、113が正常に動作していることを示す。同様にして図6(b)は、RAID管理領域119に格納されているデータ602である。データ602は、管理情報604とシステムの起動時間から構成されている。管理情報604は、ハードディスク112、113に障害が発生

していない場合の管理情報であり、ハードディスク112、113が正常に動作していることを示す。

[0021] ハードディスク112に障害が発生した場合、HDDコントローラ111は管理情報603を図7(a)に示す管理情報703に更新する。管理情報603は、ハードディスク112のRAID管理領域116に格納されているため、障害により更新することができる場合もあるし更新できない場合もある。図7(a)に示す管理情報703に記載のハードディスク112の動作状態は「(異常)」と記載してあるは、HDDコントローラ111が管理情報603を更新できた場合には、ハードディスク112の動作状態が「異常」と記載されることを意味している。そのため、HDDコントローラ111が管理情報703を更新することができない場合には、ハードディスク112の動作状態は「正常」のままである(管理情報603と同じ)。HDDコントローラ111はRAID管理領域119に格納される管理情報604を図7(b)に示す管理情報704に更新する。HDDコントローラ111は、管理情報704に記載のハードディスク112の状態を「異常」と更新し、ハードディスク113の状態を「正常」のままとする。

[0022] さらにハードディスク113に障害が発生した場合、HDDコントローラ111は管理情報703、704を図8(a)、図8(b)に示す管理情報803、804に更新する。図8(a)に示す管理情報803に記載のハードディスク112、113の動作状態は「(異常)」である。HDDコントローラ111が管理情報703を更新できた場合には、ハードディスク112、113の動作状態が「異常」と記載されることを意味している。そのため、HDDコントローラ111が管理情報703を更新することができない場合には、ハードディスク112、113の動作状態は「正常」のままである(管理情報703と同じ)。図8(b)に示す管理情報803に記載のハードディスク112の動作状態は「異常」であり、ハードディスク113の動作状態は「(異常)」である。

管理情報603、604、703、704、803、804は、ハードディスク112、113が正常に動作しているか否かを示す情報である。HDDコントローラ111は管理情報を随時更新する。

[0023] またHDDコントローラ111は、ハードディスク112、113の異常を検出する機能を有している。異常の検出の機構は具体的には例えば、診断コマンドなどをハードディス

ク112、113に送信し、それに対する応答信号がハードディスク112、113から正常応答信号が所定時間内に返信されるか否かを監視している。そしてHDDコントローラ111が所定時間内に正常応答信号を受信しない場合には、ハードディスク112、113に異常が発生したと判別する。そしてHDDコントローラ111が所定時間内に正常応答信号を受信した場合には、ハードディスク112、113に異常は発生していないと判別する。またデータの書込み又はデータの読み出しが正常に行われたか否かを監視しており、正常にデータの書込み又はデータの読み出しが行われない場合には、異常が発生したと決定する。

[0024] 図2は本実施例に係るデータの記録処理のフローチャートである。

[0025] 本実施例では、ハードディスク112に障害が発生してハードディスク112をオフラインにし、その後ハードディスク113に障害が発生してハードディスク113をオフラインにして、RAID装置100がマルチデッド状態になる場合を説明する。ハードディスクを112、113をオフラインにする順序は、ハードディスク113に障害が発生しハードディスク113をオフラインにし、その後ハードディスク112に障害が発生してハードディスク112をオフラインにする場合であってもよい。

[0026] 本実施例では、RAID装置100はミラーリング構成になっている。そのためハードディスク112、113には同一のデータが格納される。換言すれば、RAID装置100において、ハードディスク112、113が冗長構成となっている。

[0027] まずユーザはRAID装置100の電源を投入する(ステップS201)。プロセッサ106は、RAID管理領域116、119に格納されている管理情報を参照する(ステップS202)。プロセッサ106は、管理情報よりディスク112、113が正常か否かを判別する(ステップS203)。

[0028] プロセッサ106がハードディスク112、113共にオンラインでないと判別した場合には、RAID装置100はデータの記録処理を終了する(ステップS204)。プロセッサ106がハードディスク112又は、113のどちらかがオンラインであると判別した場合には、CPUモジュール101はハードディスク112のシステム領域114、又はハードディスク113のシステム領域117からOSをメモリ107に読み出しシステムを起動する(ステップS205)。そしてHDDコントローラ111はRAID管理領域116、119にシステムの起

動時刻とシステムの管理情報を書き込む(ステップS206)。RAID装置100は、RAID装置100の外部から入力されるデータをデータ領域115、118に記録しデータの更新を行う(ステップS207)。

- [0029] そしてハードディスク112に障害が発生する(ステップS208)。プロセッサ106は、ハードディスク112をオフラインにする(ステップS209)。そしてRAID装置100は、RAID装置100の外部から入力されるデータをデータ領域118に記録しデータの更新を行う(ステップS210)。ハードディスク112はオフラインのため、データ領域115はデータの更新を行うことができない(ステップS211)。
- [0030] ハードディスク113に障害が発生する(ステップS212)。HDDコントローラ111は、ハードディスク113をオフラインにする(ステップS213)。ハードディスク112、113をオフラインにしたため、RAID装置100はシステムダウンする(ステップS214)。
- [0031] 図3は本実施例に係るデータの記録復旧のフローチャートである。
- [0032] RAID装置100は、ハードディスク112、113の両方をオフラインにしてシステムダウンすると、ユーザもしくは、CPUモジュール101がシステムをリセットする(ステップS301)。そしてプロセッサ106は、RAID管理領域116、119に格納されている管理情報の参照を試みる(ステップS302)。プロセッサ106はRAID管理領域116、119の管理情報を読み出し可能か否かを判別する(ステップS303)。
- [0033] RAID管理領域116、119の管理情報を読み出し可能な場合には、プロセッサ106はRAID管理領域116、119の管理情報を参照する。
- [0034] そしてプロセッサ106は、RAID管理領域116、119の管理情報にお互いのハードディスク状態が異常として記録されるなどの矛盾があるか否かを判別する(ステップS304)。
- [0035] プロセッサ106は、RAID管理領域116、119の管理情報に矛盾がないと判別した場合には、RAID管理領域116に格納されている管理情報とRAID管理領域119に格納されている管理情報を比較し冗長構成喪失直後の状態を解析する。
- [0036] プロセッサ106はハードディスク112がRAID装置100から切り離された後、ハードディスク113が切り離されたことを判別する。そしてプロセッサ106はハードディスク113のみオンラインにして冗長構成喪失直後の状態にした後にCPUモジュール101

はシステムを起動する(ステップS305)。

- [0037] RAID管理領域116の管理情報には、HDDコントローラ111がハードディスク112へアクセスしたアクセス時刻などのログが記録されている。同様にRAID管理領域119の管理情報には、HDDコントローラ111がハードディスク113へアクセスしたアクセス時刻などのログが記録されている。プロセッサ106はこのハードディスク112、113へのアクセス時刻を比較して、ハードディスク112、ハードディスク113が切り離された順序を判別して冗長構成喪失直後の状態を解析する。またプロセッサ106はシステム領域114、117に記録されている起動時刻も参照することも可能であり、起動時刻よりハードディスク112、ハードディスク113が切り離された順序を判別して冗長構成喪失直後の状態を解析してもよい。
- [0038] RAID管理領域116、119の管理情報を読み出し不可能な場合(ステップS303 NO)や、RAID管理領域116、119の管理情報にお互いのハードディスク状態が異常として記録されるなどの矛盾があり、冗長構成喪失直後の状態を解らない場合には(ステップS304 YES)、プロセッサ106は、冗長構成喪失直後の状態を示す情報がメモリ107にある場合かを確認する。プロセッサ106はHDDコントローラ111にハードディスク112への書き込み禁止の設定を行う。またプロセッサ106は、ハードディスクをオンラインにした情報がメモリ107に記憶されているかを確認すると、記憶されていないため、メモリ107にディスク112をオンラインにした事を記憶し、ハードディスク112をオンラインにする。(ステップS306)。これによりハードディスク112をオンラインにした場合であっても、HDDコントローラ111はハードディスク112への書き込みを行わない。HDDコントローラ111はハードディスク112の全ての書き込み内容(システムの起動時刻と起動状態などを含む)をデータキャッシュ109に記録する。HDDコントローラ111はシステムの起動時刻と起動状態をデータバッファ110に記録する構成であってもよい。起動状態とは、ハードディスク112がオンラインかオフラインかを示す状態である。
- [0039] CPUモジュール101はOSがパニックや、ハングアップしシステムが正常に起動しないことを判別する(ステップS307)。ユーザもしくは、CPUモジュール101がシステムをリセットする(ステップS308)。プロセッサ106は、冗長構成喪失直後の状態を示

す情報がメモリ107にあるかを確認する。プロセッサ106はHDDコントローラ111にハードディスク113の書き込み禁止の設定を行う。プロセッサ106は直前にハードディスク112をオンラインにした情報がメモリ107に記憶されているかを確認し、ハードディスク112がオンラインになっていることを確認するため、プロセッサ106はハードディスク112をオフラインにする(ステップS309)。そしてプロセッサ106はハードディスク113をオンラインにし、メモリ107にディスク112をオフライン、ディスク113をオンラインにした事を記憶しておく。(ステップS310)。

[0040] CPUモジュール101は、OSが正常に立ち上がると、プロセッサ106に対してシステムが正常に動作した事を通知する。そしてメモリ107に格納される設定テーブルに示す書き込みフラグをすべてONにしてOSを再起動する。プロセッサ106は、冗長構成喪失直後の状態を示す情報(RAID装置100が、ハードディスク112がオフラインで、ハードディスク113がオンラインの状態であることを示す情報)をメモリ107より参照する。

[0041] そしてプロセッサ106は、ハードディスク112をオフライン、ハードディスク113をオンラインにして正常に起動する(ステップS311)。

[0042] その後、ハードディスク112は、障害の可能性が高いため新しいハードディスクに交換し、ハードディスク113からリビルドを実施する。リビルドが完了後、同様にハードディスク113を新しいディスクに交換し最初のミラー構成にし、システムを完全に復旧する。

デジーチェーン接続のハードディスクにおいてバス系に異常がある場合などは、RAID装置が不揮発性メモリなどに記録するログからだけでは、ログを記録する周期の間隔内に複数のハードディスクの切り離しが発生する事がある。これによりログからだけではハードディスクの切り離しの順序を特定することができず、適切にRAID装置を復旧することができない。本願発明によればハードディスクをオンラインにしても起動時刻などを書き込まずに起動するため、正常に起動するか否かを判別して立ち上げの順序を特定することができる。(実施例2)

本実施例は記憶装置900がデータ領域を複数のハードディスクに分割(ストライピングという)してデータのパリティを固定されたディスクに冗長化したRAID3の場合を

例に説明する。また記憶装置900は、RAID3に限定されることはなく、RAID5のようなパリティが複数のハードディスクに分散してもあってもよい。

[0043] 図9は、本実施例に係るRAID装置900のハードブロック図である。

[0044] <RAID装置900>

RAID装置900は、CPU(Central Processing Unit)モジュール901、ディスプレイコントローラ(Display Controller)902、ディスプレイ903、LANコントローラ(LAN Controller)904、RAIDコントローラ(RAID Controller)905、ハードディスク(HDD)912、913、914から構成される。CPUモジュール901は、RAID装置900を統括的に制御し、ディスプレイコントローラ902、LANコントローラ904、RAIDコントローラ905を統括的に制御する。またCPUモジュール901はメモリなどの記録媒体を含む構成であり、RAID装置900のOSをメモリ上に展開してOSを起動するものである。ディスプレイコントローラ902は、ディスプレイ903における表示制御を行う。LANコントローラ904はRAID装置900が接続するネットワークの通信制御を行う。RAID装置900は、LANコントローラ904を介してRAID装置900の外部のネットワークより、所望のデータを取得する。

[0045] <RAIDコントローラ905>

RAIDコントローラ905は、プロセッサ906、メモリ907、バスコントローラ908、データキャッシュ909、データバッファ910、HDDコントローラ911から構成されている。プロセッサ906は、HDDコントローラ911経由でハードディスク912またはハードディスク913にデータを書き込む／読み出す制御を行う。プロセッサ906はメモリ907に格納されている所定のプログラムを実行する。バスコントローラ908は、RAIDコントローラ905とCPUモジュール901でのデータのやりとりを制御する。さらにプロセッサ906はHDDコントローラ911経由でハードディスク912、913、914との接続制御を行う。RAIDコントローラ905とハードディスク912、913、914を接続することを「オンライン」、RAIDコントローラ905とハードディスク912、913、914を切断することを「オフライン」と呼ぶ。ただしオフライン、オンラインに限らず、プロセッサ906はHDDコントローラ911経由でハードディスク912、913から管理情報を読み出すことが可能である。データキャッシュ909はハードディスク912、913、914のアクセスを補助し、一度ハ

ードディスク912、913、914にアクセスするデータを一時的に保管し、保管された情報は、ディスクに直接アクセスせずに高速化する。そまたデータバッファ911は、ハードディスク912、913、914に書き込むデータを一旦、格納する記憶部である。ハードディスク912、913、914は一定時間に一定のデータ量を書き込む必要があるために、データバッファ910は設けられている。換言すれば、データキャッシュ909とデータバッファ910は、CPUモジュール901またはプロセッサ906が書き込みを指示するデータをハードディスク912、913、914に格納する前に格納する。そしてデータ量が所定の容量よりも大きくなった後、HDDコントローラ911は、データキャッシュ909経由でデータバッファ910からデータを読み出し、ハードディスク912、913、914に格納する。データキャッシュ109と、データバッファ910があることによって、RAID装置900はデータの転送をハンドリングすることができる。

[0046] ハードディスク912はシステム領域915、データ領域916、RAID管理領域917より構成されている。同様にしてハードディスク913はシステム領域918、データ領域919、RAID管理領域920より構成されている。ハードディスク914はシステム領域921、データ領域922、RAID管理領域923より構成されている。

[0047] システム領域915、データ領域916、RAID管理領域917は、ハードディスク912を分割した領域である。同様にしてシステム領域918、データ領域919、RAID管理領域920は、ハードディスク913を分割した領域であり、システム領域921、データ領域922、RAID管理領域923は、ハードディスク914をパーティション分割した領域である。そしてシステム領域915、918、921、データ領域916、919、922、RAID管理領域917、920、923の容量は可変な量であって、ハードディスク912、913、914のデータ容量の範囲内で可変である。

[0048] RAID3は、データのパリティ情報を格納するハードディスクをストライプアレイに付加した構成である。これによりハードディスクに障害が発生した場合、RAID3はパリティ情報よりデータの再構築を行うことができる。本実施例において、ストライピングアレイはハードディスク912、913である。パリティ情報を格納するハードディスクはハードディスク914である。

[0049] システム領域915には、RAID装置900におけるOS、アプリケーションプログラムな

どのストライプデータ1201が格納されている。同様にしてシステム領域918には、RAID装置900におけるOS、アプリケーションプログラムなどのストライプデータ1202が格納されている。ストライプデータ1201、1202は、RAID装置900におけるOS、アプリケーションプログラムなどのデータを分散したデータである。ストライプデータ1201、1202の分散比率等は、ユーザが設定することができる。そしてシステム領域921には、ストライプデータ1201とストライプデータ1202の排他的論理和であるパリティデータ9211が格納されている。

[0050] データ領域916、919にはRAID装置900のデータが分散され、それぞれにストライプデータ9161、9191格納されている。そしてデータ領域921にはデータ領域916に格納されるストライプデータ9161とデータ領域919に格納されるストライプデータ9191との排他的論理和であるパリティデータ9221が格納されている。ここでデータとはユーザの個人情報などであって、RAID装置100を制御するプログラムなどとは異なる情報である。

[0051] RAID管理領域917、920、923にはRAID装置900の管理情報が分散されずそれぞれに独立して管理データ917、920、923に格納されている。管理情報とは、ハードディスク912、913、914が正常か否かを示す情報であり、プロセッサ906がHDDコントローラ911経由で更新を行う。

[0052] 図12(a)は、本実施例に係るシステム領域915に格納されているストライプデータ1201である。ストライプデータ1201は、RAID装置900において起動するOS1205から構成されている。ストライプデータ1201は、RAID装置900におけるOS、アプリケーションプログラムから構成されるデータを分散したデータであるため、OS1205、アプリケーションプログラム1207は、これらOS、アプリケーションプログラム、設定テーブルの一部分である。同様にして図12(b)は、本実施例に係るシステム領域918に格納されているストライプデータ1202である。ストライプデータ1202は、RAID装置900において起動するOS1206、アプリケーションプログラム1208から構成されている。OS1205とOS1206の和がRAID装置900のOSである。同様にしてアプリケーションプログラム1207とアプリケーションプログラム1208の和がRAID装置900のアプリケーションプログラムである。本実施例において、ストライプデータ1201、120

2は共に、OS、アプリケーションプログラムの一部分を有するが、分散の仕方はこれに限定されることはない。例えばシステム領域915にはRAID装置900のOSを格納し、システム領域918はRAID装置900のアプリケーションプログラムを格納する構成であってもよい。

[0053] 図19に記載の設定テーブル2001はハードディスク912、ハードディスク913、ハードディスク914、データキャッシュ909、データバッファ110へのデータの書込み／読み出しを決定する書き込みフラグのオン／オフの一部を示すテーブルである。初期設定では設定テーブル2001の書き込みフラグはすべてオンになっている。設定テーブル2001はメモリ907に格納されている。

[0054] 本実施例において、RAID装置900はパラレルアクセスアレイ構成になっているため、OS1205とOS1206は異なる内容を示すデータである。同様にして、アプリケーションプログラム1207と1208も異なる内容を示すデータである。

[0055] さらにパリティデータ1301はストライプデータ1201とストライプデータ1202の排他的論理和のデータである。プロセッサ906は、パリティデータ1301を算出する。パリティデータ1301は、そのため例えばハードディスク912に障害が発生し、ストライプデータ1201を読み出すことができない場合、パリティデータ1301とストライプデータ1202からストライプデータ1201を復元することができる。同様にしてハードディスク913に障害が発生し、ストライプデータ1202を読み出すことができない場合、パリティデータ1301とストライプデータ1201からストライプデータ1202を復元することができる。

[0056] 本実施例において、BIOS (Basic Input/Output System) 設定などにおいて、ユーザはデータキャッシュ109、データバッファ110、ハードディスク912、913、914の書込みを有効/無効にするフラグのオン／オフを設定する。ハードディスク912、913、914への書込みフラグをオフにすると、設定テーブル2001が更新される。データキャッシュ909へデータの書込みを許可する書き込みフラグをオンのままにして変更せず、データバッファ110への書き込みを許可する書き込みフラグもオンのままにする。ユーザ若しくはCPUモジュール901がRAID装置900をリブートしたとき、ハードディスク912、913、914の書込みフラグをオフに設定することによって、ハード

ディスク912、913、914へ起動に伴う初期データの書き込みを禁止することができる。初期データはOSの起動時刻、起動状態などを含み、システムの起動に伴う設定データやログデータである。

[0057] 図14(a)は、RAID管理領域917に格納されている管理データ1401である。管理データ1401は、管理情報1403とシステムの起動時間1405から構成されている。管理情報1403は、ハードディスク912、913、914に障害が発生していない場合の管理情報であり、ハードディスク912、913、914が正常に動作していることを示す。同様にして図14(b)は、RAID管理領域920に格納されている管理データ1402である。管理データ1402は、管理情報1404とシステムの起動時間1406から構成されている。管理情報1404は、ハードディスク912、913、914に障害が発生していない場合の管理情報であり、ハードディスク912、913、914が正常に動作していることを示す。同様にして図14(c)は、RAID管理領域923に格納されている管理データ1407である。管理データ1407は、管理情報1408とシステムの起動時間1409から構成されている。管理情報1408は、ハードディスク912、913、914に障害が発生していない場合の管理情報であり、ハードディスク912、913、914が正常に動作していることを示す。

[0058] ハードディスク912に障害が発生した場合、プロセッサ906は、HDDコントローラ911経由で、管理情報1403を図15(a)に示す管理情報1603に更新する。管理情報1603は、ハードディスク912のRAID管理領域917に格納されているため、障害により更新することができる場合もあるし更新できない場合もある。図15(a)に示す管理情報1603に記載のハードディスク912の動作状態は「(異常)」と記載してあるは、HDDコントローラ911が管理情報1403を更新できた場合には、ハードディスク912の動作状態が「異常」と記載されることを意味している。プロセッサ906は、ハードディスク913、914の状態を「正常」のままとする。そのため、プロセッサ906は、HDDコントローラ911経由で、管理情報1403を更新することができない場合には、ハードディスク912の動作状態は「正常」のままである。プロセッサ906はRAID管理領域920に格納される管理情報1404を図15(b)に示す管理情報1604に更新する。プロセッサ906は、管理情報1404に記載のハードディスク912の状態を「異常」と更新し、ハー

ドディスク913、914の状態を「正常」のままとする。プロセッサ906はRAID管理領域923に格納される管理情報1407を図15(c)に示す管理情報1607に更新する。プロセッサ906は、管理情報1407に記載のハードディスク912の状態を「異常」と更新し、ハードディスク913、914の状態を「正常」のままとする。

[0059] さらにハードディスク913に障害が発生した場合、プロセッサ906は管理情報1603、1604、1607を図16(a)、図16(b)、図16(c)に示す管理情報1703、1704、1708に更新する。図16(a)に示す管理情報1603に記載のハードディスク912、913の動作状態は「(異常)」である。プロセッサ906が管理情報1603を更新できた場合には、ハードディスク912、913の動作状態が「異常」と記載されることを意味している。そのため、プロセッサ906が管理情報1603を更新することができない場合には、ハードディスク912、913の動作状態は「正常」のままである(管理情報1603と同じ)。図16(b)に示す管理情報1703に記載のハードディスク912の動作状態は「異常」であり、ハードディスク913の動作状態は「(異常)」であり、ハードディスク914の動作状態は「正常」のままである。図16(c)に示す管理情報1707に記載のハードディスク912、913の動作状態は「異常」であり、ハードディスク914の動作状態は「正常」のままである。

[0060] 管理情報1403、1404、1407、1603、1604、1407、1703、1704、1407は、ハードディスク912、913、914が正常に動作しているか否かを示す情報である。プロセッサ906は、管理情報を随時更新する。また管理情報を比較して矛盾がある場合あり、矛盾とは、管理情報に記載されているハードディスク912、913、914の「正常」、「異常」の一致が取れないことである。

またHDDコントローラ911は、ハードディスク912、913、914の異常を検出する機能を有している。異常の検出の機構は例えば、HDDコントローラ911が診断コマンドなどの信号をハードディスク912、913、914に送信し、それに対する正常応答信号がハードディスク912、913、914から所定時間内に返信されるか否かを監視している。そしてHDDコントローラ911が所定時間内に正常応答信号を受信しない場合には、ハードディスク912、913、914に異常が発生したと判別する。そしてHDDコントローラ911が所定時間内に応答信号を受信した場合には、ハードディスク912、913、91

4に異常は発生していないと判別する。またデータの書込み又はデータの読み出しが正常に行われたか否かを監視しており、正常にデータの書込み又はデータの読み出しが行われない場合には、異常が発生したと決定する。図10は本実施例に係るデータの記録処理のフローチャートである。

- [0061] 本実施例では、ハードディスク912に障害が発生しハードディスク912をオフラインにし、その後ハードディスク913に障害が発生してハードディスク913をオフラインにして、RAID装置900がマルチデッド状態になる場合を説明する。ハードディスク912、913をオフラインにする順序は、ハードディスク913をオフラインにし、その後ハードディスク912をオフラインにする場合であってもよい。
- [0062] まずユーザはRAID装置900の電源を投入する(ステップS1001)。プロセッサ906は、RAID管理領域917、920、923に格納されている管理情報を参照する(ステップS1002)。
- [0063] CPUモジュール901はシステム領域915、918、921からOSをメモリ907に読み出し、システムを起動する(ステップS1003)。そしてプロセッサ906は、HDDコントローラ911経由で、RAID管理領域917、920、923にシステムの起動時刻とシステムの管理情報を書き込み、管理データを更新する(ステップS1004)。RAID装置900は、RAID装置900の外部から入力されるデータをデータ領域916、918、922に記録しデータの更新を行う(ステップS1005)。プロセッサ906は、データ領域916のストライプデータとデータ領域919のストライプデータの排他的論理和を計算し、パリティデータを算出する(ステップS1006)。そしてプロセッサ906は、HDDコントローラ911経由で、パリティデータをデータ領域922に格納する(ステップS1007)。
- [0064] そしてハードディスク912に障害が発生する(ステップS1008)。プロセッサ906は、HDDコントローラ911経由で、ハードディスク912をオフラインにする(ステップS1009)。そしてRAID装置900は、RAID装置900の外部から入力されるデータをデータ領域919に記録しデータの更新を行う(ステップS1010)。プロセッサ906は、HDDコントローラ911経由で、データ領域922のパリティデータの更新を行う(ステップS1011)ハードディスク912はオフラインのため、データ領域916はデータの更新を行うことができない(ステップS1012)。

- [0065] ハードディスク913に障害が発生する(ステップS1013)。プロセッサ906は、HDDコントローラ911経由で、ハードディスク913をオフラインにする(ステップS1014)。ハードディスク912、913をオフラインにしたため、RAID装置900はシステムダウンする(ステップS1015)。
- [0066] 図11は本実施例に係るデータの記録復旧のフローチャートである。
- [0067] RAID装置900は、ハードディスク912、913の両方を切り離してシステムダウンすると、ユーザもしくは、CPUモジュール101がシステムをリセットする(ステップS1101)。そしてプロセッサ906は、RAID管理領域917、920、923に格納されている管理情報の参照を試みる(ステップS1102)。プロセッサ906はRAID管理領域917、920、923の管理情報を読み出し可能か否かを判別する(ステップS1103)。
- [0068] RAID管理領域917、920、923の管理情報を読み出し可能な場合には、プロセッサ906はRAID管理領域917、920、923の管理情報を参照する。
- [0069] そしてプロセッサ906は、RAID管理領域917、920、923の管理情報にハードディスク状態が異常として記録されるなどの矛盾があるか否かを判別する(ステップS1104)。
- [0070] プロセッサ906は、RAID管理領域917、920、923の管理情報に矛盾がないと判別した場合には、RAID管理領域917、920、923に格納されている管理情報を比較し、冗長構成喪失直後の状態を解析する。
- [0071] プロセッサ906はハードディスク912をオフラインにした後に、ハードディスク913をオフラインにしたことを判別する。そしてプロセッサ906はハードディスク913のみオンラインにして(クリティカル状態にして)、ユーザ若しくは、CPUモジュール901はシステムを起動する(ステップS1105)。
- [0072] RAID管理領域917には、HDDコントローラ911がハードディスク912へアクセスしたアクセス時刻などのログが記録されている。同様にRAID管理領域920には、HDDコントローラ911がハードディスク913へアクセスしたアクセス時刻などのログが記録されている。同様にRAID管理領域923には、HDDコントローラ911がハードディスク914へアクセスしたアクセス時刻などのログが記録されている。プロセッサ906はこのHDDコントローラ911がハードディスク912、913、914へアクセスしたアクセス時

刻を比較して、ハードディスク912、ハードディスク913をオフラインにした順序を判別し、冗長構成喪失直後の状態を解析する。

[0073] またプロセッサ906はシステム領域915、918に記録されている起動時刻、及びシステム領域921のパリティデータを参照することも可能である。起動時刻よりハードディスク912、ハードディスク913がオフラインになった順序を判別する構成であってもよい。

[0074] RAID管理領域917、920、923の管理情報を読み出し不可能な場合(ステップS1103 NO)や、RAID管理領域917、920、923の管理情報にハードディスク状態が異常として記録されるなどの矛盾があり、冗長構成喪失直後の状態を解らない場合には(ステップS1104 YES)、プロセッサ906は、冗長構成喪失直後の状態を示す情報がメモリ907にある場合かを確認する。プロセッサ906はHDDコントローラ911にハードディスク912への書き込み禁止の設定を行う。またプロセッサ906は、ハードディスク912をオンラインにした情報がメモリ907に記憶されているかを確認すると、記憶されていないため、メモリ907にディスク912をオンラインにした事を記憶し、ハードディスク912をオンラインにする(ステップS1106)。HDDコントローラ911がハードディスク912への書き込みは行わないように、ユーザは予めメモリ907に格納されている設定テーブルのハードディスク912、913、914の書き込みフラグをオフにしておく。そのためハードディスク912をオンラインにした場合であっても、HDDコントローラ911はハードディスク912への書き込みを行わない。HDDコントローラ911はハードディスク912、914の全ての書き込み内容(システムの起動時刻と起動状態などを含む)をデータキャッシュ909に記録する。HDDコントローラ911は912、914の全ての書き込み内容をデータバッファ910に記録する構成であってもよい。ハードディスク912の起動状態とは、ハードディスク912がオンラインかオフラインかを示す状態である。

[0075] CPUモジュール901はOSがパニックや、ハングアップしシステムが正常に起動しないことを判別する(ステップS1107)。ユーザもしくは、CPUモジュール901がシステムをリセットする(ステップS1108)。プロセッサ906はハードディスク912をオフラインにする(ステップS1109)。そしてプロセッサ906はハードディスク913をオンラインにし、メモリ907にディスク912をオフライン、ディスク913、914をオンラインにした事

を記憶しておく。(ステップS1110)。

[0076] CPUモジュール901は、OSが正常に立ち上がると、プロセッサ906に対してシステムが正常に動作した事を通知する。そしてメモリ907に格納される設定テーブルに示す書き込みフラグをすべてONにしてOSを再起動する。プロセッサ906は、冗長構成喪失直後の状態を示す情報(RAID装置900が、ハードディスク912がオフラインで、ハードディスク913、914がオンラインの状態であることを示す情報)をメモリ907より参照する。

[0077] そしてプロセッサ906は、ハードディスク912をオフライン、ハードディスク913、914をオンラインにして正常に起動する(ステップS1111)。またこのとき、RAID管理領域920、923に格納されている管理情報に示すハードディスク913の動作状態を「正常」に更新する。

[0078] その後、ハードディスク912は、障害の可能性が高いため新しいハードディスクに交換し、ハードディスク913、914からリビルドを実施する。リビルドが完了後、同様にハードディスク913を新しいディスクに交換し最初のRAID3構成にし、システムを完全に復旧する。

本実施例におけるRAID装置900では、ハードディスク912、913、914のうち複数のハードディスクに障害が発生した場合であっても、正しい順序でハードディスクをオンラインにしてシステムを正常に起動することができる。ハードディスクのオンラインにする順序を特定して、ハードディスク912、913、914へのデータの誤書き込みを防止できるため、RAID装置900はハードディスク912、913、914のデータを救済することができる。

本発明に係る判別手段が行う処理は、プロセッサが行う処理に含まれる。本発明における起動手段が行う処理は、CPUモジュールが行う処理に含まれる。本発明における格納手段が行う処理は、データキャッシュ、データバッファが行う処理に含まれる。本発明に係る稼働手段、特定手段が行う処理は、CPU、プロセッサが行う処理に含まれる。本発明に係る書き込み手段が行う処理は、HDDコントローラが行う処理に含まれる。また本実施例において、データを記録する記録媒体はハードディスク912、913、914であるが、これに限られることはなく、揮発性記録媒体であるメモリや不

揮発性記録媒体のフラッシュメモリなどであってもよい。

産業上の利用可能性

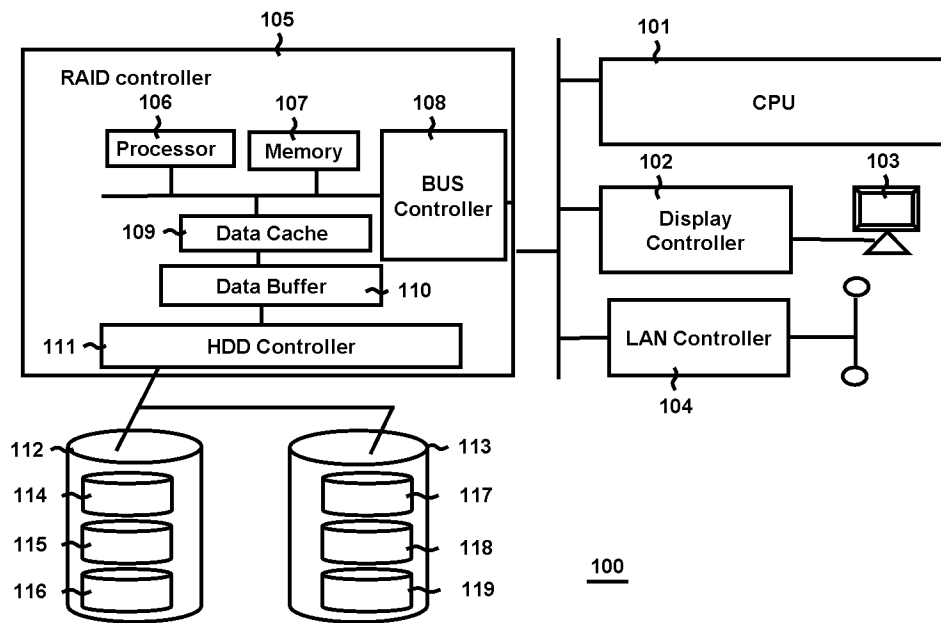
[0079] 本発明による記憶装置は、記録されたデータの復旧を行うものである。したがって、本発明による記憶装置は、障害の発生したディスクに記録されたデータを取得するうえで極めて有用である。

請求の範囲

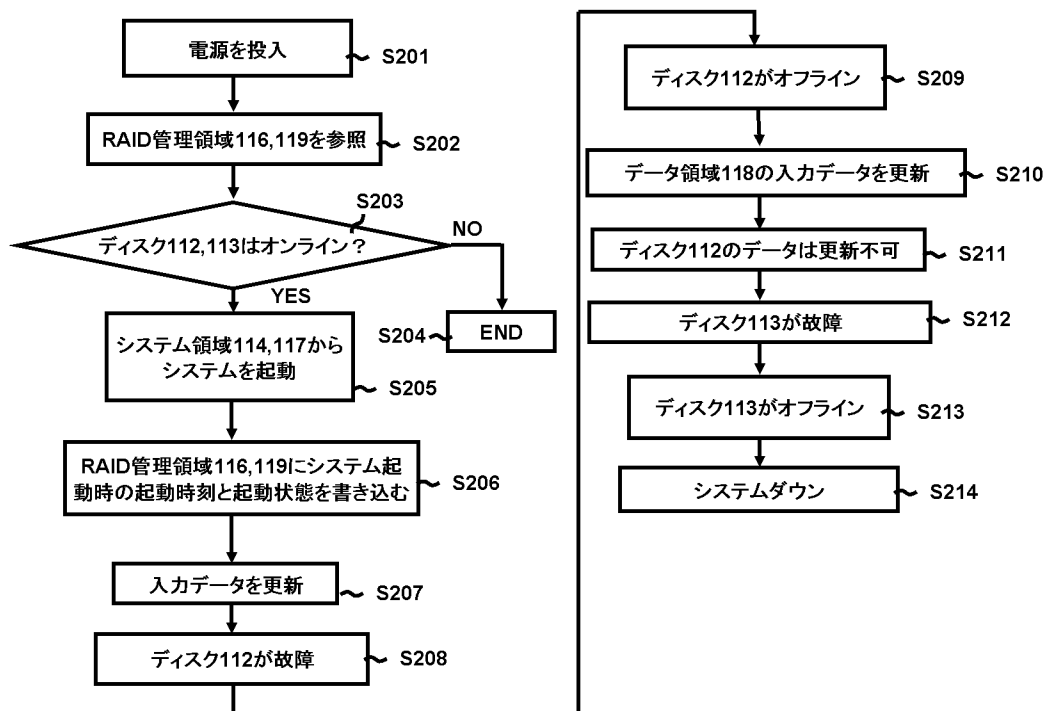
- [1] 複数の記録媒体を有し、障害を検出した記録媒体を休止する記憶装置において、
休止した記録媒体が複数であるか否かを判別する判別手段と、
複数の記録媒体の休止を判別した該記憶装置を再起動する起動手段と、
該記憶装置の起動に伴う初期データを格納する格納手段と、
休止した記録媒体のいずれか一つを稼動する稼動手段と、
該稼動した記録媒体に該初期データの書込みを禁止し、稼動する記憶装置の稼動順序を特定する特定手段と、
該特定手段において、稼動順序が正しいと判別した記録媒体に対して、該格納手段に格納した該初期データを記録する書込み手段と、
からなることを特徴とする記憶装置。
- [2] 請求項1に記載の記憶装置において、
該特定手段は、稼動した記録媒体に記録される履歴情報を参照して該稼動順序を特定することを特徴とする記憶装置。
- [3] 請求項1に記載の記憶装置において、
該特定手段は、該記憶装置の動作設定を参照して該稼動順序を特定することを特徴とする記憶装置。
- [4] 請求項1に記載の記憶装置において、
該特定手段は、該記憶装置が有するオペレーションシステムの動作状態を参照して該稼動順序を特定することを特徴とする記憶装置。
- [5] 請求項1に記載の記憶装置において、
該複数の記録媒体に同一のデータを格納することを特徴とする記憶装置。
- [6] 請求項1に記載の記憶装置において、
該複数の記録媒体にデータを分散して記録し、さらに分散した分散データの排他的論理和を格納することを特徴とする記憶装置。
- [7] 請求項1に記載の記憶装置において、
該格納手段は、該記憶装置の動作を効率化するキャッシュメモリであることを特徴とする記憶装置。

- [8] 請求項1に記載の記憶装置において、
該格納手段は、該記録媒体に書き込むデータを一時的に格納するデータバッファであることを特徴とする記憶装置。
- [9] 請求項1に記載の記憶装置において、
該書込み手段は、該稼動した記録媒体が有するフラグを変更して該初期データを記録することを特徴とする記憶装置。
- [10] 複数の記録媒体を有し、障害を検出した記録媒体を休止する記憶装置が記録するデータを復旧する記録復旧方法において、
休止した記録媒体が複数であるか否かを判別する判別手順と、
複数の記録媒体の休止を判別した該記憶装置を再起動する起動手段と、
該記憶装置の起動に伴う初期データを格納する格納手順と、
休止した記録媒体のいずれか一つを稼動する稼動手順と、
該稼動した記録媒体に該初期データの書込みを禁止し、稼動する記憶装置の稼動順序を特定する特定手順と、
稼動順序が正しいと判別した記録媒体に対して、該格納手段に格納した該初期データを記録する書込み手順と、
からなることを特徴とする記録復旧方法。
- [11] 複数の記録媒体を有し、障害を検出した記録媒体を休止する記憶装置が記録するデータを復旧する記録復旧プログラムにおいて、
休止した記録媒体が複数であるか否かを判別する判別ステップと、
複数の記録媒体の休止を判別した該記憶装置を再起動する起動ステップと、
該記憶装置の起動に伴う初期データを格納する格納ステップと、
休止した記録媒体のいずれか一つを稼動する稼動ステップと、
該稼動した記録媒体に該初期データの書込みを禁止し、稼動する記憶装置の稼動順序を特定する特定ステップと、
稼動順序が正しいと判別した記録媒体に対して、該格納手段に格納した該初期データを記録する書込みステップと、
をコンピュータに実行させることを特徴とする記録復旧プログラム。

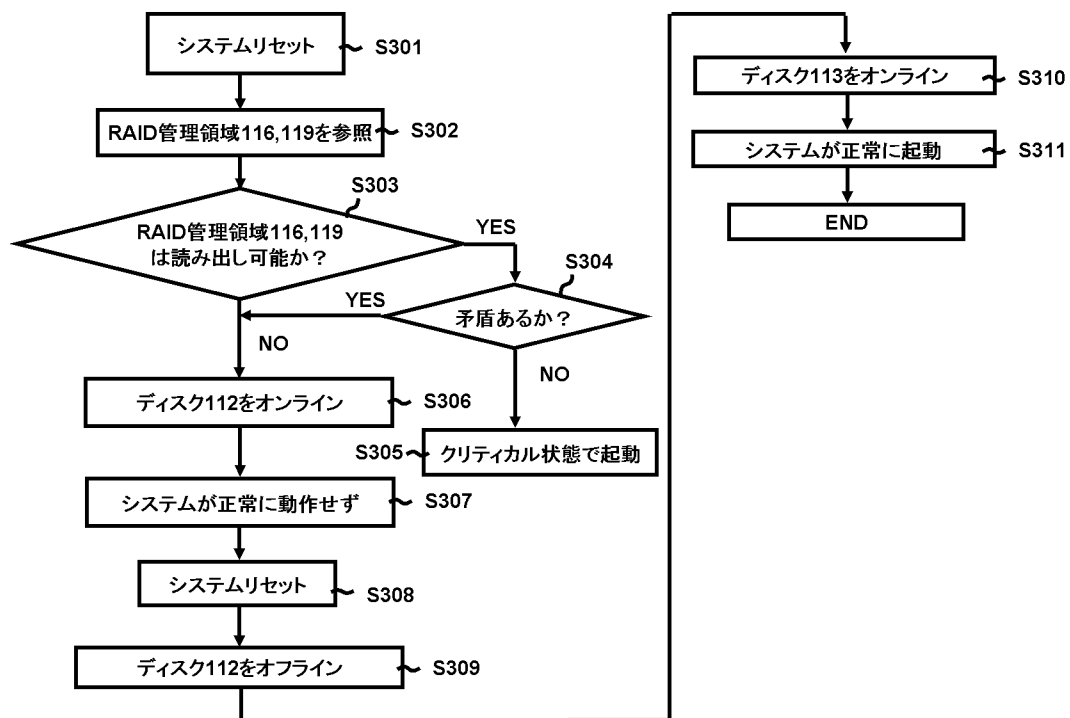
[図1]



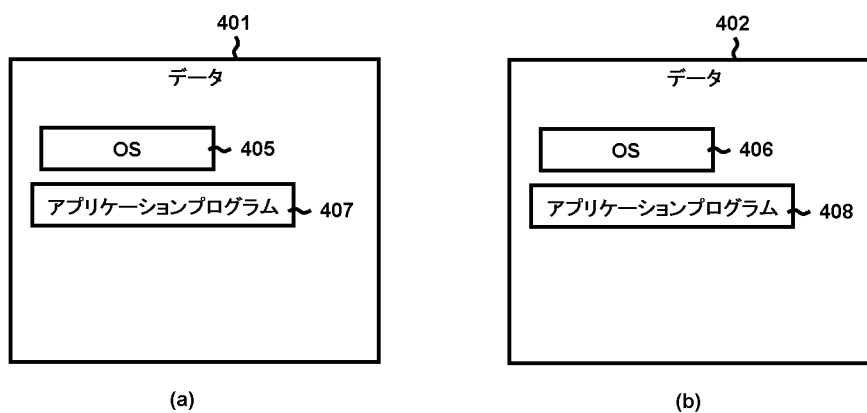
[図2]



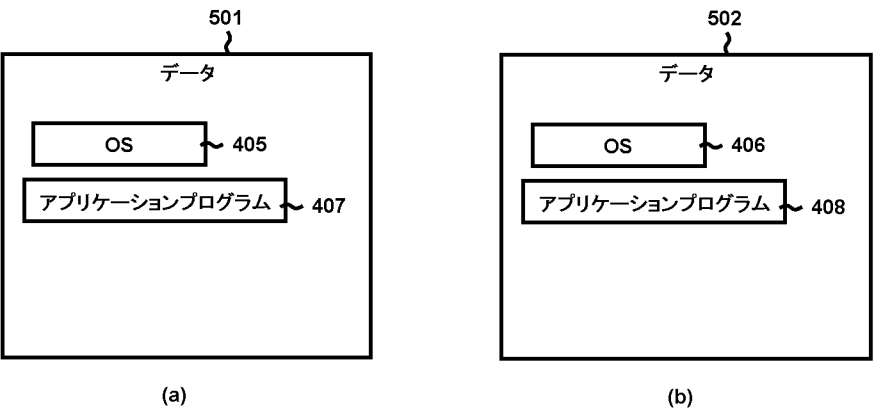
[図3]



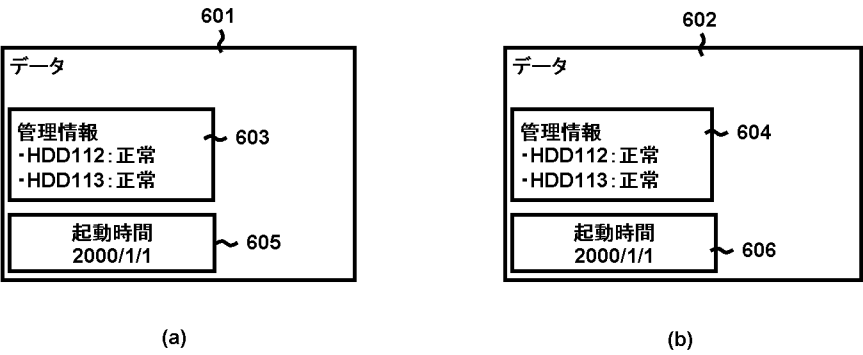
[図4]



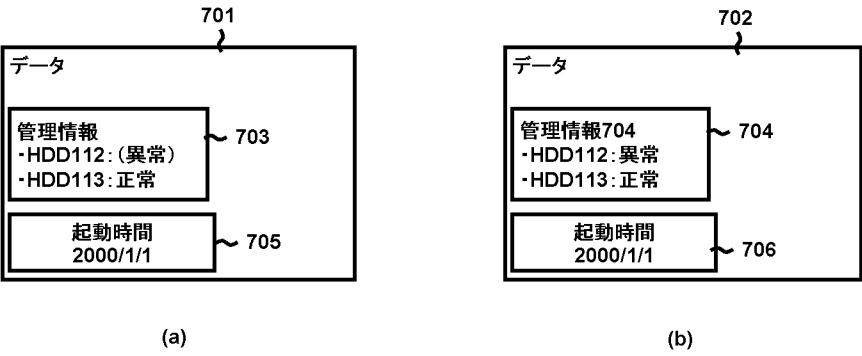
[図5]



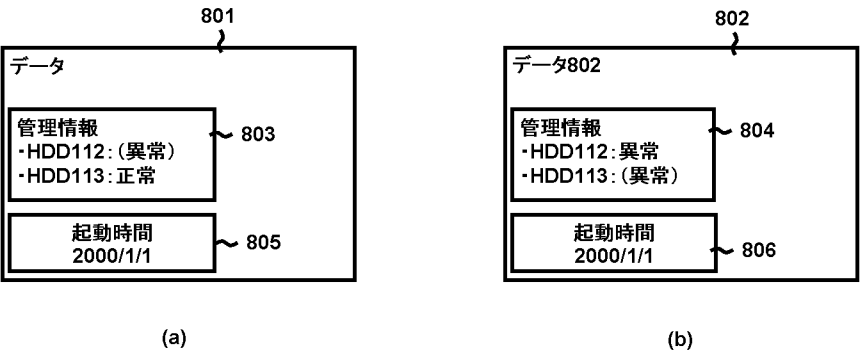
[図6]



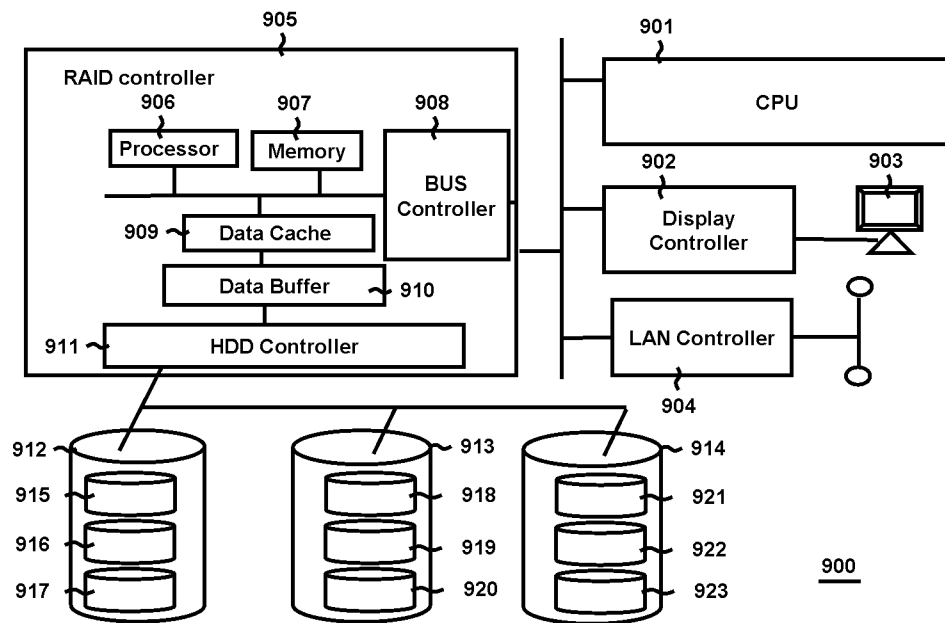
[図7]



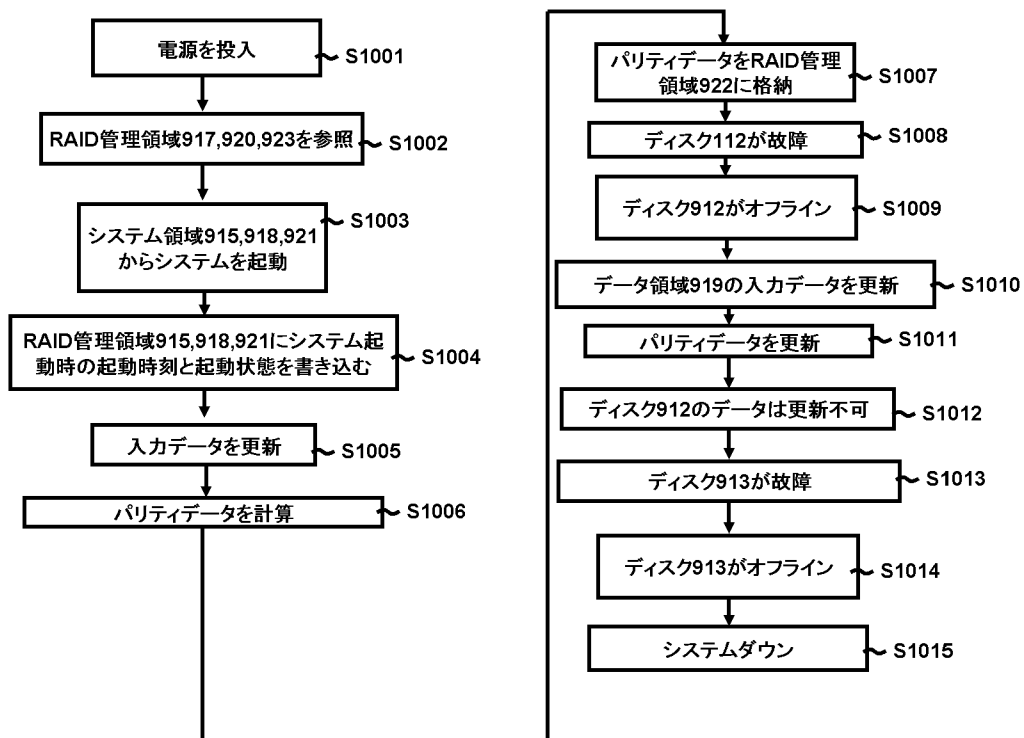
[図8]



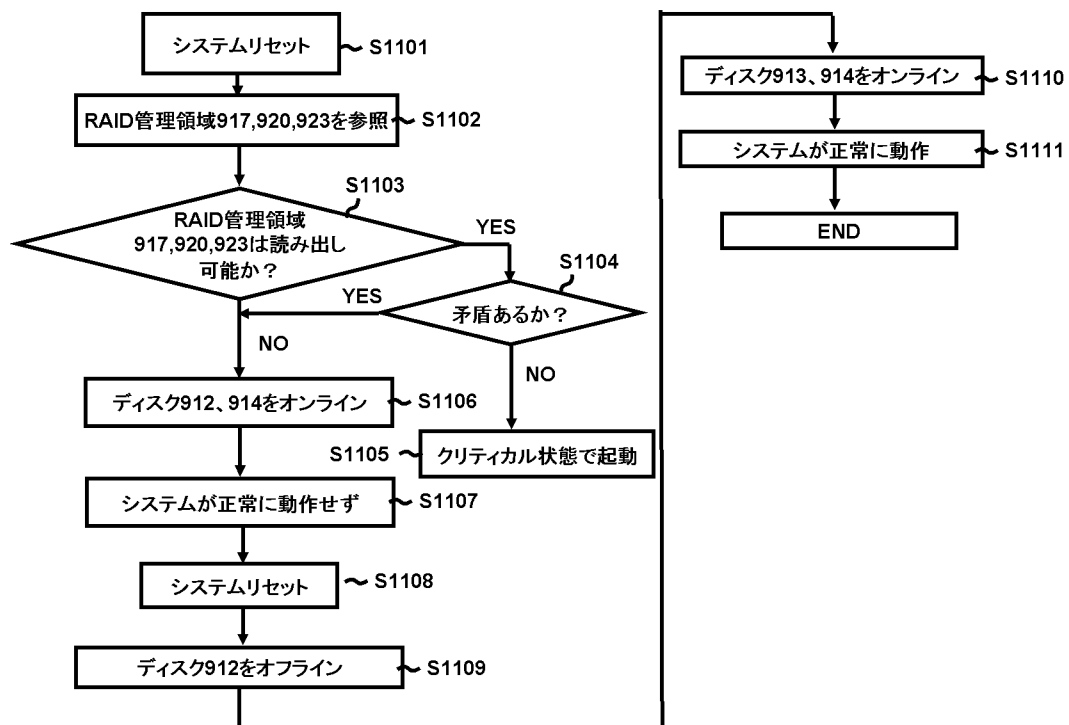
[図9]



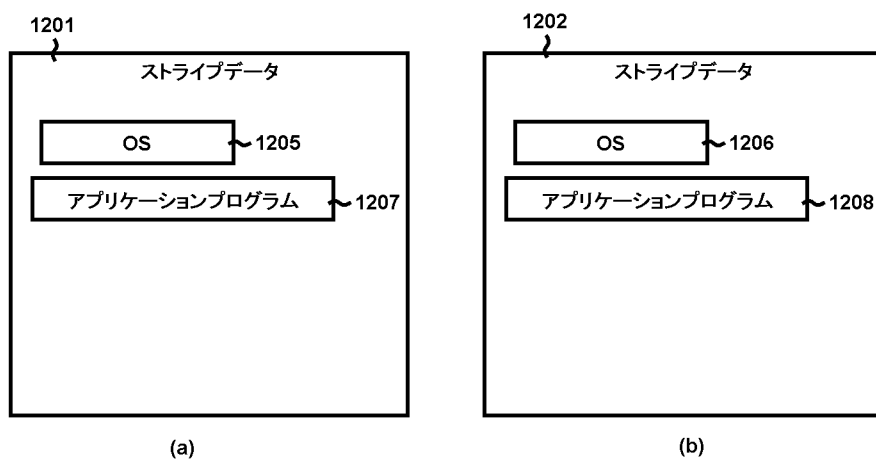
[図10]



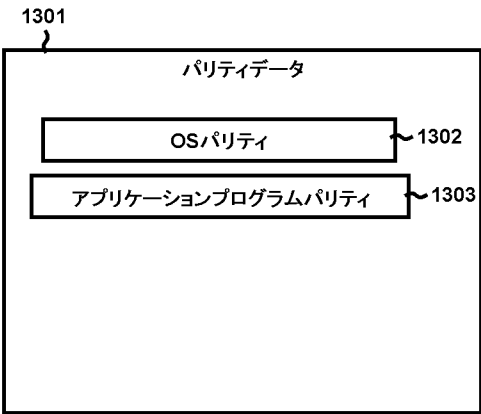
[図11]



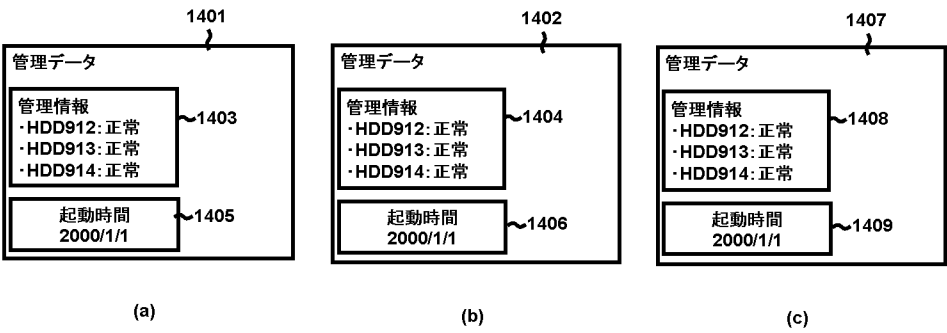
[図12]



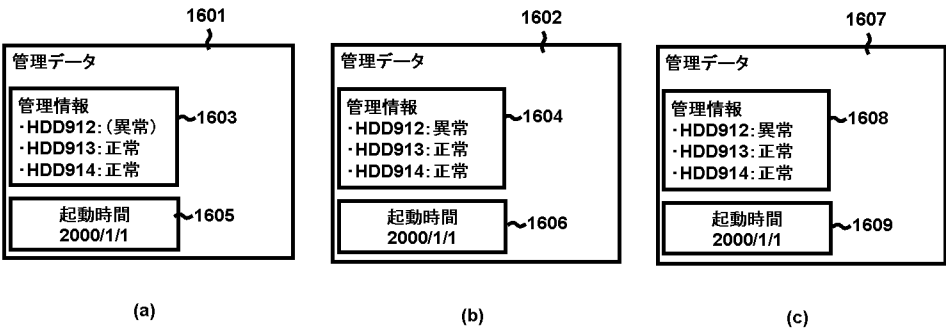
[図13]



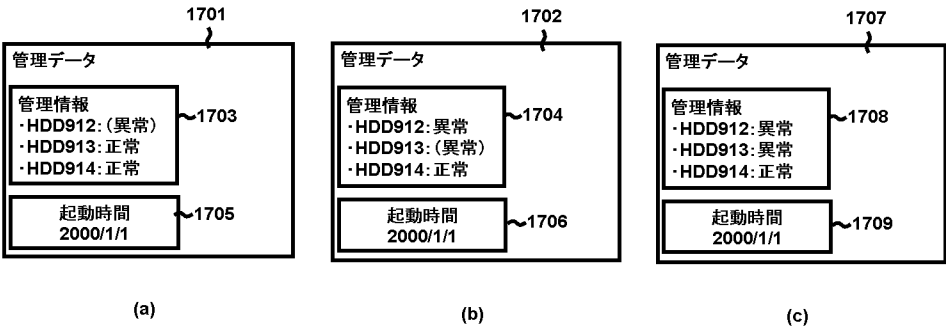
[図14]



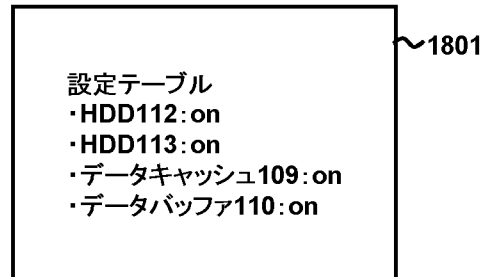
[図15]



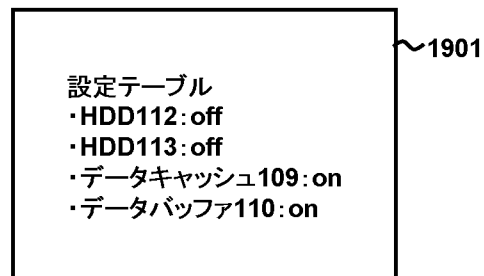
[図16]



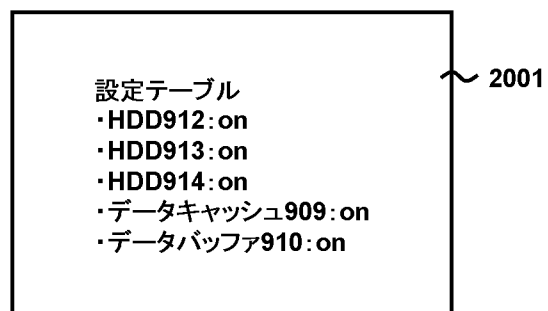
[図17]



[図18]



[図19]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2007/051195

A. CLASSIFICATION OF SUBJECT MATTER

G06F3/06 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F3/06

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2007
Kokai Jitsuyo Shinan Koho	1971-2007	Toroku Jitsuyo Shinan Koho	1994-2007

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 3-229331 A (Fujitsu Ltd.), 11 October, 1991 (11.10.91), Page 2, lower left column, line 15 to page 3, lower left column, line 8 (Family: none)	1-11
A	JP 58-62899 A (Toshiba Corp.), 14 April, 1983 (14.04.83), Full text; all drawings (Family: none)	1-11
A	JP 2005-250827 A (Clarion Co., Ltd.), 15 September, 2005 (15.09.05), Par. No. [0057] (Family: none)	1-11

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
20 April, 2007 (20.04.07)

Date of mailing of the international search report
01 May, 2007 (01.05.07)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2007/051195

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 7-56694 A (Fuji Xerox Co., Ltd.), 03 March, 1995 (03.03.95), Par. Nos. [0024] to [0043] (Family: none)	1-11
A	JP 11-328823 A (NEC Engineering Kabushiki Kaisha), 30 November, 1999 (30.11.99), Full text; all drawings (Family: none)	1-11

A. 発明の属する分野の分類（国際特許分類（IPC））

Int.Cl. G06F3/06(2006.01)i

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

Int.Cl. G06F3/06

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2007年
日本国実用新案登録公報	1996-2007年
日本国登録実用新案公報	1994-2007年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	JP 3-229331 A（富士通株式会社）1991.10.11, 第2頁左下欄第15行目－第3頁左下欄第8行目 （ファミリーなし）	1-11
A	JP 58-62899 A（株式会社東芝）1983.04.14, 全文、全図（ファミリーなし）	1-11
A	JP 2005-250827 A（クラリオン株式会社）2005.09.15, 段落【0057】（ファミリーなし）	1-11

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日

20.04.2007

国際調査報告の発送日

01.05.2007

国際調査機関の名称及びあて先

日本国特許庁（ISA/J P）
郵便番号100-8915
東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

須藤 竜也

5 I

3801

電話番号 03-3581-1101 内線 3565

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	JP 7-56694 A (富士ゼロックス株式会社) 1 9 9 5 . 0 3 . 0 3 , 段落【0 0 2 4】－【0 0 4 3】 (ファミリーなし)	1－1 1
A	JP 11-328823 A (日本電気エンジニアリング株式会社) 1 9 9 9 . 1 1 . 3 0 , 全文、全図 (ファミリーなし)	1－1 1