

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4203034号
(P4203034)

(45) 発行日 平成20年12月24日(2008.12.24)

(24) 登録日 平成20年10月17日(2008.10.17)

(51) Int.Cl. F I
G 0 6 F 3/06 (2006.01) G O 6 F 3/06 3 O 4 B
 G O 6 F 3/06 5 4 O

請求項の数 8 (全 13 頁)

(21) 出願番号	特願2005-86358 (P2005-86358)	(73) 特許権者	000003078 株式会社東芝
(22) 出願日	平成17年3月24日(2005.3.24)		東京都港区芝浦一丁目1番1号
(65) 公開番号	特開2006-268502 (P2006-268502A)	(73) 特許権者	301063496 東芝ソリューション株式会社
(43) 公開日	平成18年10月5日(2006.10.5)		東京都港区芝浦一丁目1番1号
審査請求日	平成17年3月24日(2005.3.24)	(74) 代理人	100058479 弁理士 鈴江 武彦
		(74) 代理人	100091351 弁理士 河野 哲
		(74) 代理人	100088683 弁理士 中村 誠
		(74) 代理人	100108855 弁理士 蔵田 昌俊

最終頁に続く

(54) 【発明の名称】 アレイコントローラ、メディアエラー修復方法及びプログラム

(57) 【特許請求の範囲】

【請求項1】

論理ディスクを構成する複数のディスクドライブへのアクセスを制御することによりホストから要求されたデータと当該データの冗長データとを前記複数のディスクドライブに分散して配置するアレイコントローラにおいて、

前記論理ディスクを構成する前記複数のディスクドライブの各々の故障を予知するディスクドライブ故障予知手段と、

前記論理ディスクを構成する前記複数のディスクドライブのうち前記ディスクドライブ故障予知手段によって故障が予知されたディスクドライブを除く全てのディスクドライブの内容を検査してメディアエラーの有無を確認するための健全性確認処理を実行する健全性確認手段と、

前記健全性確認手段によって、メディアエラーが検出された場合に、前記論理ディスクを構成する前記複数のディスクドライブのうち当該メディアエラーが発生しているディスクドライブを除く全てのディスクドライブのデータまたは冗長データを利用して当該メディアエラーが発生している箇所のデータの欠損を修復するメディアエラー修復手段と、

前記故障が予知されたディスクドライブのデータをスペアのディスクドライブにコピーするためのデータコピー処理を前記健全性確認手段による前記健全性確認処理が進行するアドレスの方向とは逆方向に実行するデータコピー手段と、

前記データコピー手段による前記データコピー処理の完了後に、前記故障が予知されたディスクドライブを前記スペアのディスクドライブに代替させて、前記故障が予知された

10

20

ディスクドライブを前記論理ディスクから切り離す論理ディスク復元手段とを具備し、

前記健全性確認手段は、前記データコピー手段による前記データコピー処理が完了した領域に対応するアドレスまで前記健全性確認処理を実行し終えた段階で当該健全性確認処理を終了する

ことを特徴とするアレイコントローラ。

【請求項 2】

前記データコピー手段は、通常は前記故障が予知されたディスクドライブのデータを前記スペアのディスクドライブにコピーし、前記故障が予知されたディスクドライブのデータを正常に読み出すことができなかつた場合には当該データを前記論理ディスクを構成する前記複数のディスクドライブのうち前記故障が予知されたディスクドライブを除く全てのディスクドライブのデータまたは冗長データから復元して、その復元されたデータを前記スペアのディスクドライブにコピーすることを特徴とする請求項 1記載のアレイコントローラ。

10

【請求項 3】

前記データコピー手段は、前記故障が予知されたディスクドライブのデータを、前記論理ディスクを構成する前記複数のディスクドライブのうち前記故障が予知されたディスクドライブを除く全てのディスクドライブのデータまたは冗長データから復元して、その復元されたデータを前記スペアのディスクドライブにコピーすることを特徴とする請求項 1記載のアレイコントローラ。

【請求項 4】

前記データコピー手段は、前記データコピー処理を前記健全性確認手段による前記健全性確認処理の開始時に開始することを特徴とする請求項 1記載のアレイコントローラ。

20

【請求項 5】

前記ディスクドライブ故障予知手段は、前記論理ディスクを構成する前記複数のディスクドライブの各々が標準的に有する、当該ディスクドライブの信頼性悪化に係る状態を当該ディスクドライブ自身が監視して分析し、その結果を当該ディスクドライブを利用する装置に報告する機能を利用して、前記複数のディスクドライブの各々の故障を予知することを特徴とする請求項 1記載のアレイコントローラ。

【請求項 6】

前記ディスクドライブ故障予知手段は、前記論理ディスクを構成する前記複数のディスクドライブの各々に対するアクセスでエラーリトライまたは回復可能なエラーが発生した回数をカウントし、当該複数のディスクドライブの各々のカウント値に基づいて当該複数のディスクドライブの各々の故障を予知することを特徴とする請求項 1記載のアレイコントローラ。

30

【請求項 7】

論理ディスクを構成する複数のディスクドライブへのアクセスを制御することによりホストから要求されたデータと当該データの冗長データとを前記複数のディスクドライブに分散して配置するアレイコントローラに適用されるメディアエラー修復方法であって、

前記論理ディスクを構成する前記複数のディスクドライブの各々の故障を予知するステップと、

40

前記複数のディスクドライブのうちのいずれかのディスクドライブの故障が予知された場合、前記論理ディスクを構成する前記複数のディスクドライブのうち前記故障が予知されたディスクドライブを除く全てのディスクドライブの内容を検査してメディアエラーの有無を確認するための健全性確認処理を実行するステップと、

前記健全性確認処理によってメディアエラーが検出された場合に、前記論理ディスクを構成する前記複数のディスクドライブのうち当該メディアエラーが発生しているディスクドライブを除く全てのディスクドライブのデータまたは冗長データを利用して当該メディアエラーが発生している箇所のデータの欠損を修復するステップと、

前記故障が予知されたディスクドライブのデータをスペアのディスクドライブにコピーするためのデータコピー処理を前記健全性確認処理が進行するアドレスの方向とは逆方向

50

に実行するステップと、

前記データコピー処理が完了した領域に対応するアドレスまで前記健全性確認処理を実行し終えた段階で当該健全性確認処理を終了させるステップと、

前記データコピー処理の完了後に、前記故障が予知されたディスクドライブを前記スペアのディスクドライブに代替させて、前記故障が予知されたディスクドライブを前記論理ディスクから切り離すステップと

を具備することを特徴とするメディアエラー修復方法。

【請求項 8】

論理ディスクを構成する複数のディスクドライブへのアクセスを制御することによりホストから要求されたデータと当該データの冗長データとを前記複数のディスクドライブに分散して配置するアレイコントローラに実行させるためのプログラムであって、

前記アレイコントローラに、

前記論理ディスクを構成する前記複数のディスクドライブの各々の故障を予知するステップと、

前記複数のディスクドライブのうちのいずれかのディスクドライブの故障が予知された場合、前記論理ディスクを構成する前記複数のディスクドライブのうち前記故障が予知されたディスクドライブを除く全てのディスクドライブの内容を検査してメディアエラーの有無を確認するための健全性確認処理を実行するステップと、

前記健全性確認処理によってメディアエラーが検出された場合に、前記論理ディスクを構成する前記複数のディスクドライブのうち当該メディアエラーが発生しているディスクドライブを除く全てのディスクドライブのデータまたは冗長データを利用して当該メディアエラーが発生している箇所のデータの欠損を修復するステップとと、

前記故障が予知されたディスクドライブのデータをスペアのディスクドライブにコピーするためのデータコピー処理を前記健全性確認処理が進行するアドレスの方向とは逆方向に実行するステップと、

前記データコピー処理が完了した領域に対応するアドレスまで前記健全性確認処理を実行し終えた段階で当該健全性確認処理を終了させるステップと、

前記データコピー処理の完了後に、前記故障が予知されたディスクドライブを前記スペアのディスクドライブに代替させて、前記故障が予知されたディスクドライブを前記論理ディスクから切り離すステップと

を実行させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、複数のディスクドライブから構成される冗長性を持つディスクアレイを制御するアレイコントローラに係り、特にディスクドライブのメディアエラーを迅速に解消するのに好適な、アレイコントローラ、メディアエラー修復方法及びプログラムを提供することにある。

【背景技術】

【0002】

冗長データを持つことによりデータの信頼性を向上させる技術として、複数のディスクドライブ(HDD)を用いて構成される冗長性を持つディスクアレイ(冗長化ディスクアレイ)、つまりRAID(Redundant Array of Inexpensive Disks、またはRedundant Array of Independent Disks)が知られている。RAIDには幾つかのレベル(RAIDレベル)が定義されており、RAID1(ミラーリング)やRAID5(パリティ付きストライピング)などが知られている。いずれも複数のHDDを用いて構成される冗長化ディスクアレイにデータ及び冗長データを配置することにより、いずれかのHDDが故障してもデータの復元を可能にする技術である。

【0003】

ディスクアレイと当該ディスクアレイを制御するアレイコントローラとから構成される

10

20

30

40

50

装置は、ディスクアレイ装置と呼ばれる。ディスクアレイは、ディスクアレイ装置を外部記憶装置として利用するホスト（ホストコンピュータ）からは、1つの記憶領域を有する1つの論理ディスク（論理ユニット）として認識される。このため、ディスクアレイは、論理ディスク（論理ユニット）と呼ばれることもある。

【0004】

ディスクアレイ装置では、ディスクアレイ（論理ディスク）を構成する複数のHDDのいずれかが故障した場合、その故障したHDD（故障HDD）を別の正常なHDD（スペアHDD）に交換するのが一般的である（例えば、特許文献1参照）。アレイコントローラは、ディスクアレイを構成している複数のHDDのうち、故障HDDを除くHDDのデータにより、当該ディスクアレイにおけるデータの冗長性を利用して、故障HDDのデータを復元する。復元されたデータは、交換されたHDDに格納される。このようにして故障HDDのデータが、交換されたHDDに復元される。するとディスクアレイ装置は、HDDの故障発生前と同様に動作を継続することができる。

10

【特許文献1】特開平8-190460号公報（段落0020乃至0027）

【発明の開示】

【発明が解決しようとする課題】

【0005】

上記したように従来技術においては、ディスクアレイを構成する複数のHDDのいずれかが故障した場合、その故障したHDD（故障HDD）のデータを、残りのHDDのデータから復元することができる。

20

【0006】

ところが、故障HDD以外の、ディスクアレイを構成する複数のHDDのいずれかにメディアエラーが存在することがある。メディアエラーとは、HDDのディスク媒体（ディスクメディア）に起因して、データを読み出したり書き込むことができないエラーを指す。もし、メディアエラーが存在する場合、故障HDDのデータを復元するのに必要なデータまたは冗長データを、当該メディアエラーが存在するHDDから読み出すことができなくなるおそれがある。このような場合、メディアエラーが存在する領域に対応する故障HDDのデータを復元できなくなる。

【0007】

例えばHDD#1及び#2から構成されるRAID1レベルの論理ディスクにおいて、HDD#1が故障した場合を考える。この場合、故障したHDD#1を新たなHDD#3に交換する。この状態で、アレイコントローラは、正常なHDD#2からデータをリードし、新たなHDD#3へそのリードデータのライトを行うことで、データを復元する。ところが、HDD#2からのデータリードで当該HDD#2にメディアエラーが発生すると、その部分のデータを読み出すことができず、HDD#3へデータをライトすることができない。この結果、メディアエラーの発生部分に格納されていたデータは失われ、復元は失敗となる。

30

このように従来の技術では、故障したHDD（ディスクドライブ）のデータを復元する際、正常なHDD（ディスクドライブ）のメディアエラーにより復元失敗となる場合がある。

40

【0008】

本発明は上記事情を考慮してなされたものでその目的は、正常なディスクドライブに存在するメディアエラーを事前に、或いはできるだけ迅速に解消することにより、復元失敗の発生を防止できるアレイコントローラ、メディアエラー修復方法及びプログラムを提供することにある。

【課題を解決するための手段】

【0009】

本発明の1つの観点によれば、論理ディスクを構成する複数のディスクドライブへのアクセスを制御することによりホストから要求されたデータと当該データの冗長データとを上記複数のディスクドライブに分散して配置するアレイコントローラが提供される。この

50

アレイコントローラは、上記論理ディスクを構成する複数のディスクドライブの各々の故障を予知するディスクドライブ故障予知手段と、上記論理ディスクを構成する複数のディスクドライブのうち上記ディスクドライブ故障予知手段によって故障が予知されたディスクドライブを除く全てのディスクドライブの内容を検査してメディアエラーの有無を確認するための健全性確認処理を実行する健全性確認手段と、この健全性確認手段によって、メディアエラーが検出された場合に、上記論理ディスクを構成する複数のディスクドライブのうち当該メディアエラーが発生しているディスクドライブを除く全てのディスクドライブのデータまたは冗長データを利用して当該メディアエラーが発生している箇所のデータの欠損を修復するメディアエラー修復手段とから構成される。

【0010】

10

このような構成においては、論理ディスクを構成する複数のディスクドライブのいずれかが近い将来に故障することがディスクドライブ故障予知手段によって予知された場合に、当該複数のディスクドライブのうち、故障が予知されたディスクドライブ以外の全てのディスクドライブの内容が健全性確認手段によって検査される。この検査で、データが正常に読み出せないメディアエラーが検出された場合、上記論理ディスクを構成する複数のディスクドライブのうち当該メディアエラーが発生しているディスクドライブを除く全てのディスクドライブ（つまり故障が予測されたディスクドライブを含むディスクドライブ）のデータまたは冗長データを利用して当該メディアエラーが発生している箇所のデータの欠損が修復される。

【0011】

20

このように上記の構成においては、論理ディスクを構成する複数のディスクドライブの中から近い将来に故障する可能性のあるディスクドライブを予知して、当該故障する可能性のあるディスクドライブが実際に故障する前に、つまり当該故障する可能性のあるディスクドライブのデータまたは冗長データを正常に読み出すことができ論理ディスクの冗長性が保たれている間に、論理ディスクを構成する他のディスクドライブのメディアエラーが発生している箇所が、当該故障する可能性のあるディスクドライブのデータまたは冗長データをも利用して修復される。これにより、その後、上記故障する可能性のあるディスクドライブが実際に故障したとしても、論理ディスクを構成する残りのディスクドライブのデータまたは冗長データを利用して、当該故障したディスクドライブのデータを復元できる。

30

【0012】

ここで、上記健全性確認手段による健全性確認処理の完了後に、上記故障が予知されたディスクドライブのデータをスペアのディスクドライブにコピーするためのデータコピー処理を実行するデータコピー手段と、このデータコピー手段によるデータコピー処理の完了後に、上記故障が予知されたディスクドライブを上記スペアのディスクドライブに代替させて、上記故障が予知されたディスクドライブを論理ディスクから切り離す論理ディスク復元手段とを追加すると良い。このようにすると、近い将来故障する可能性のあるディスクドライブが実際に故障する前に、当該ディスクドライブを確実にスペアのディスクドライブに代替させることができる。また、データコピー手段によるデータコピー処理の実行中に、故障が予想されていたディスクドライブが実際に故障してしまった場合でも、論理ディスクを構成する他のディスクドライブのデータの健全性が既に確認されていることで、より安全にスペアディスクへのデータコピーを行うことができる。

40

【0013】

また、上記のデータコピー手段に代えて、上記データコピー処理を上記健全性確認手段による健全性確認処理が進行するアドレスの方向とは逆方向に実行するデータコピー手段を用い、このデータコピー手段によるデータコピー処理が完了した領域に対応するアドレスまで上記健全性確認処理が実行された段階で当該健全性確認処理を終了する構成とすると良い。このようにすると、上記健全性確認手段による健全性確認処理の完了を待ってデータコピー手段によるデータコピー処理を開始する必要がないため、故障する可能性のあるディスクドライブを検出してからデータコピー手段によるデータコピー処理が完了する

50

までの時間を短縮でき、しかも健全性確認手段による健全性確認処理も無駄なく効率的に行える。この効果は、データコピー手段によるデータコピー処理を、健全性確認手段による健全性確認処理の開始時に開始する場合に最も高くなる。

【発明の効果】

【0014】

本発明によれば、故障する可能性のあるディスクドライブが実際に故障する前に、論理ディスクを構成する他のディスクドライブのメディアエラーが発生している箇所を、当該故障する可能性のあるディスクドライブのデータまたは冗長データをも利用して修復することにより、当該故障する可能性のあるディスクドライブが実際に故障したとしても、論理ディスクを構成する残りのディスクドライブのデータまたは冗長データを利用して、当該故障したディスクドライブのデータを復元することができ、メディアエラーによるデータ修復不能を回避することができる。

10

【発明を実施するための最良の形態】

【0015】

以下、本発明の実施の形態につき図面を参照して説明する。

図1は本発明の一実施形態に係るディスクアレイ装置の構成を示すブロック図である。図1のディスクアレイ装置は、主として論理ディスク(ディスクアレイ)10と当該論理ディスク10を制御するコントローラ(アレイコントローラ)とから構成される。論理ディスク10は当該ディスク10の信頼性の向上に必要な冗長データを保持するために、複数のディスクドライブ、例えば3台の磁気ディスクドライブ(HDD)11-0(#0)~11-2(#2)から構成されるものとする。つまり論理ディスク10は冗長性を持つ論理ユニットである。この冗長性のレベルとして、RAID1, RAID10, RAID3, RAID5などが知られている。ここでは、論理ディスク10がRAID5を適用するディスクアレイであるものとする。RAID5では、データは複数のHDDに分散して書き込まれると共に、そのデータのパリティデータが冗長データとして別のHDDに書き込まれる。また、RAID5では、冗長データ(パリティデータ)の書き込み先となるHDDは固定されていない。したがって、論理ディスク10がRAID5を適用するものとする、当該論理ディスク10を構成するHDD11-0~11-2は、いずれもデータ並びに冗長データ(パリティデータ)の格納用に用いられ、冗長データ(パリティデータ)は各HDD11-0~11-2に分散して格納される。

20

30

【0016】

図1のディスクアレイ装置は更に、論理ディスク10を構成するHDD11-0~11-2のいずれかが故障した場合に、その故障したHDDに代えて用いられるスペアのHDD(SHDD)11-3(#3)を有する。このHDD11-3を含むHDD群、即ちHDD11-0~11-3は、例えばSCSIインタフェースによりアレイコントローラ20と接続されている。

【0017】

アレイコントローラ20は、構成管理部21と、アクセス制御部22と、論理ディスク復元部23と、HDD故障予知部24と、HDD健全性確認部25と、メディアエラー修復部26と、データコピー部27とを有する。

40

【0018】

構成管理部21は、主として論理ディスク10の構成を管理する。アクセス制御部22は、図1のディスクアレイ装置を外部記憶装置として利用するホスト(ホストコンピュータ)からの読み出し/書き込み要求を受けて論理ディスク10に対するアクセスを制御する。ここではアクセス制御部22は、ホストからの読み出し/書き込み要求を、実際にデータを読み出し/書き込みすべき個々のHDDに対する読み出し/書き込み要求に変換し、その変換された読み出し/書き込み要求を該当するHDDに発行する。例えば、ホストからのデータDの書き込み要求の場合、アクセス制御部22は当該データDをデータD1及びD2に分割し、例えばHDD11-0に対してはデータD1の書き込み要求を、HDD11-1に対してはデータD2の書き込み要求を、そしてHDD11-2に対してはデータD

50

1及びD2の排他的論理和データ(パリティデータ)、つまりデータD1及びD2の冗長データの書き込み要求を、それぞれ発行する。論理ディスク復元部23は、論理ディスク10を構成するHDD群(HDD11-0~11-2)のいずれかが故障した場合に、その故障したHDD(故障HDD)をスペアのHDD(SHDD)11-4で代替する。このとき論理ディスク復元部23は、論理ディスク10内の正常なHDDのデータから故障HDDのデータを復元してHD11-4に格納する。このような構成管理部21、アクセス制御部22及び論理ディスク復元部23の各機能は、アレイドコントローラ20が一般に有する従来からよく知られている機能である。

【0019】

HDD故障予知部24は、論理ディスク10を構成するHDD11-0~11-2の故障を予知する。この故障予知の手法については後述する。HDD健全性確認部25は、論理ディスク10を構成するHDD11-0~11-2のうち、HDD故障予知部24によって故障が予知されたHDD以外の全てのHDDの内容を検査してメディアエラーの有無を検出するためのHDD健全性確認処理を、例えば一定のサイズのデータブロック(ここでは64KBのデータブロック)を単位に実行する。HDD健全性確認部25は、健全性の確認で、メディアエラーが発生したデータブロックを検出した場合、そのデータブロックをメディアエラー修復部26により修復させる。データコピー部27は、HDD故障予知部24によって故障が予知されたHDDのデータを例えば一定のサイズのデータブロック(ここでは64KBのデータブロック)を単位にスペアのHDD11-3にコピーする。

【0020】

次に、図1のディスクアレイ装置におけるアレイドコントローラ20の動作について、(1)故障HDDの予知、(2)故障HDDの予知に基づく処理、(3)HDD健全性確認処理に分けて順に説明する。

【0021】

(1)故障HDDの予知

まず、アレイドコントローラ20のHDD故障予知部24による、論理ディスク10を構成するHDD11-i(i=0,1,2)の故障の予知(予測)について説明する。

【0022】

近年のHDDは、SMART(Self-Monitoring, Analysis and Reporting Technology)と呼ばれる機能を有しているものが多い。このSMART機能は、例えばHDDが自身の信頼性悪化に関係する状態を監視して分析し、その結果をホストに報告する機能である。このSMART機能を、HDD11-iも有しているものとする、HDD故障予知部24は、このHDD11-iのSMART機能、つまりHDD11-iが持つ当該HDD11-i自身の故障を予知する機能を利用することで、簡単に当該HDD11-iの故障を予知することができる。

【0023】

さて、HDD11-iの故障の危険性に関係する事象として、

a) アレイドコントローラ20のアクセス制御部22からHDD11-iへのリクエストでエラーを起こし、アレイドコントローラ20でエラーリトライが発生した場合

b) HDD11-iへのリクエストで回復可能なエラー(HDD11-iの備えるエラー回復機能により回復されたエラー)が発生した場合

【0024】

明らかなように、HDD11-iに近い将来故障する危険性がある場合、HDD11-iへのリクエストで上記a)またはb)の事象が発生する回数は増加する。そこでHDD故障予知部24は、HDD11-iへのリクエストで上記a)またはb)の事象が発生する回数をカウントし、当該事象が予め定められた時間内に予め定められた回数を超えて発生した場合に、当該HDD11-iの故障を予知する。なお、上記a)またはb)の事象が、単に予め定められた回数を超えて発生した場合に、HDD11-iの故障を予知するようにしても構わない。

10

20

30

40

50

【 0 0 2 5 】

(2) 故障 HDD の予知に基づく処理、

次に、HDD故障予知部 2 4 によって HDD 1 1 -i の故障が予知された場合の処理について、図 2 のフローチャートを参照して説明する。

【 0 0 2 6 】

今、HDD故障予知部 2 4 が、HDD 1 1 -0 ~ 1 1 -2 のうちのいずれかの HDD 1 1 -i、例えば HDD 1 1 -0 の故障を予知 (予測) したものとす。この場合、HDD故障予知部 2 4 は、HDD 1 1 -0 の故障を予知したことを HDD 健全性確認部 2 5 及びデータコピー部 2 7 に通知して、当該 HDD 健全性確認部 2 5 及びデータコピー部 2 7 をそれぞれ起動する (ステップ S 1 , S 2) 。

10

【 0 0 2 7 】

すると、HDD健全性確認部 2 5 は HDD 健全性確認処理を実行して、その時点において論理ディスク 1 0 を構成している HDD 1 1 -0 ~ 1 1 -2 のうち、HDD故障予知部 2 4 によって故障が予知された HDD 1 1 -0 を除く全ての HDD (つまり HDD 1 1 -1 及び 1 1 -2) に実装されているディスク媒体 (メディア) の健全性を確認する。HDD健全性確認部 2 5 は、この HDD 健全性確認処理でメディアエラーが発生している箇所を検出した場合、メディアエラー修復部 2 6 により、その箇所のデータを修復させる。この修復には、論理ディスク 1 0 を構成する HDD 1 1 -0 ~ 1 1 -2 のうち、メディアエラーが発生している HDD を除く全ての HDD (つまり、故障が予知された HDD 1 1 -0 を含む HDD) のデータまたは冗長データが用いられる。

20

【 0 0 2 8 】

この HDD 健全性確認部 2 5 及びメディアエラー修復部 2 6 の動作は、HDD故障予知部 2 4 により HDD 1 1 -0 の故障が予知されたことで、当該 HDD 1 1 -0 が近い将来に本当に故障するおそれがあることを考慮して行われる。つまり、HDD 1 1 -0 が実際に故障して当該 HDD 1 1 -0 のメディアにアクセスできなくなる前に、故障が予知されていない他の HDD 1 1 -1 及び 1 1 -2 のメディアエラーを検出し、そのメディアエラーの箇所のデータを、論理ディスク 1 0 を構成する HDD 1 1 -0 ~ 1 1 -2 のうち、メディアエラーが発生している HDD を除く全ての HDD のデータまたは冗長データを用いて修復する。

【 0 0 2 9 】

もし、故障が予知された HDD 1 1 -0 が近い将来に本当に故障した場合には、上述のような HDD 1 1 -1 及び 1 1 -2 のメディアエラーの箇所の修復ができなくなってしまう。これに対して本実施形態では、HDD 1 1 -0 の故障が予知された段階で、故障が予知されていない他の HDD 1 1 -1 及び 1 1 -2 のメディアエラー箇所を修復するため、その修復に当該 HDD 1 1 -0 (つまり、この段階では未だ故障していない HDD 1 1 -0) のデータまたは冗長データを利用できる。

30

【 0 0 3 0 】

一方、データコピー部 2 7 は、HDD故障予知部 2 4 によって故障が予知された HDD 1 1 -0 のデータをスペアの HDD 1 1 -3 に一定のサイズのデータブロックを単位にコピーするためのデータコピー処理を実行する。このデータコピー処理で、HDD 1 1 -0 からデータが読み出せない箇所 (つまりメディアエラーが発生している箇所) が存在した場合、データコピー部 2 7 は、論理ディスク 1 0 を構成する他の HDD 1 1 -1 及び 1 1 -2 のデータを用いて HDD 1 1 -0 から読み出すべきデータを復元して、HDD 1 1 -3 にコピーする。このことは、データコピーが行われたデータブロックは、データの冗長性を保つ健全性が確認されたブロックであることを示す。もし故障が予知された HDD 1 1 -0 が、データコピー部 2 7 によるデータコピー処理の期間に実際に故障した場合にも、データコピー部 2 7 は上記と同様に、論理ディスク 1 0 を構成する他の HDD 1 1 -1 及び 1 1 -2 のデータを用いて HDD 1 1 -0 から読み出すべきデータを復元して、HDD 1 1 -3 にコピーする。なお、データコピー部 2 7 によるデータコピーの期間に、ホストからデータ書き込みが要求された結果、HDD 1 1 -0 にデータまたは冗長データを書き込む場合、アクセス制御部 2 2 は HDD 1 1 -3 にも同一のデータまたは冗長データを書き込む。

40

50

【 0 0 3 1 】

ここで、データコピー部 2 7 によるデータコピー処理と HDD 健全性確認部 2 5 による HDD 健全性確認処理とは、処理対象 HDD を HDD 1 1 - j で表すと、当該 HDD 1 1 - j のアドレス 0 から最大アドレスの方向に実行されても、最大アドレスからアドレス 0 の方向に実行されても構わない。しかし本実施形態では、処理の効率化のために、データコピー部 2 7 による処理の方向と HDD 健全性確認部 2 5 による処理の方向とは逆方向となっている。ここでは、HDD 健全性確認部 2 5 による処理（HDD 健全性確認処理）は、図 3（a）において矢印 3 1 で示されるように、HDD 1 1 - j の最大アドレスからアドレス 0 の方向に実行される。これに対して、データコピー部 2 7 による処理（データコピー処理）は、図 3（a）において矢印 3 2 で示されるように、HDD 1 1 - j のアドレス 0 から最大アドレスの方向に実行される。この処理方向の違いによる効果については後述する。なお本実施形態において、データコピー部 2 7 による処理の対象となる HDD 1 1 - j は HDD 1 1 - 0 及び 1 1 - 3 であって、HDD 健全性確認部 2 5 による処理の対象となる HDD 1 1 - j は HDD 1 1 - 1 及び 1 1 - 2 であるというように、両 HDD 1 1 - j は相違する。しかし、図 3 では、便宜的に 1 つの HDD 1 1 - j で代表させてある。

10

【 0 0 3 2 】

さて、HDD 故障予知部 2 4 は、HDD 健全性確認部 2 5 及びデータコピー部 2 7 を起動すると、データコピー部 2 7 によるデータコピー処理の完了を待つ（ステップ S 3）。もし、データコピー部 2 7 によるデータコピー処理が完了すると、HDD 故障予知部 2 4 は論理ディスク復元部 2 3 を起動する（ステップ S 4）。すると論理ディスク復元部 2 3 は、HDD 故障予知部 2 4 によって故障が予知された HDD 1 1 - 0 をスペアの HDD 1 1 - 3 で代替させ、当該 HDD 1 1 - 0 を論理ディスク 1 0 から切り離す。このときスペアの HDD 1 1 - 3 には、上記ステップ S 2 のデータコピー部 2 7 によるデータコピー処理で HDD 1 1 - 0 のデータがコピーされている。したがって論理ディスク復元部 2 3 は、HDD 1 1 - 0 が実際に故障したために当該 HDD 1 1 - 0 をスペアの HDD 1 1 - 3 で代替させる場合と異なり、HDD 1 1 - 0 のデータを論理ディスク 1 0 の他の HDD 1 1 - 1 及び 1 1 - 2 のデータまたは冗長データで復元する処理を行う必要はない。論理ディスク 1 0 内の HDD 1 1 - 0 がスペアの HDD 1 1 - 3 で代替されて、当該 HDD 1 1 - 1 が HDD 1 1 - 0 から切り離されると、構成管理部 2 1 は論理ディスク 1 0 の構成管理情報を更新する。

20

【 0 0 3 3 】

(3) HDD 健全性確認処理

次に、HDD 健全性確認部 2 5 による HDD 健全性確認処理の詳細について、図 4 のフローチャートを参照して説明する。

30

【 0 0 3 4 】

HDD 健全性確認部 2 5 は、健全性確認の対象とすべきデータブロック（論理ブロック）を指定するアドレス（論理ブロックアドレス）から、HDD 1 1 - j（ここでは HDD 1 1 - 1 及び 1 1 - 2）の全領域をスキャン（つまり健全性を確認）したかを判定する（ステップ S 1 1）。もし、スキャン（健全性を確認）すべき領域が残っているならば、HDD 健全性確認部 2 5 は、健全性確認の対象とすべきアドレスが、故障が予知された HDD 1 1 - 0 からスペアの HDD 1 1 - 3 へのデータコピーが完了している領域に含まれているかを判定する（ステップ S 1 2）。

40

【 0 0 3 5 】

もし、健全性確認の対象とすべきアドレスが、HDD 1 1 - 0 からスペアの HDD 1 1 - 3 へのデータコピーが完了している領域に含まれていないならば、HDD 健全性確認部 2 5 は、当該アドレスの指定する領域（データブロック）を検査して当該領域の健全性を確認するための例えばリードコマンドを HDD 1 1 - 1 及び 1 1 - 2 に発行する（ステップ S 1 3）。HDD 健全性確認部 2 5 は、このリードコマンドに対する HDD 1 1 - 1 及び 1 1 - 2 からの応答に基づき、当該 HDD 1 1 - 1 または 1 1 - 2 でメディアエラーが発生したかを判定する（ステップ S 1 4）。もし、メディアエラーが発生していないならば、HDD 健全性確認部 2 5 は健全性確認の対象とすべきアドレスを次のデータブロックを指定するよう

50

更新してステップS 1 1に戻る。

【0036】

これに対し、HDD 1 1-1またはHDD 1 1-2でメディアエラーが発生しているならば、HDD健全性確認部25はそのメディアエラーが発生している箇所(データブロック)をメディアエラー修復部26により修復させる(ステップS 15)。今、HDD 1 1-2でメディアエラーが発生しているものとする、そのメディアエラー発生箇所の修復は、当該メディアエラー発生箇所(データブロック)に対応するHDD 1 1-0及び1 1-1のデータまたは冗長データで、当該メディアエラー発生箇所のデータを復元し、その復元されたデータをメディアエラーが発生したデータブロック(アドレス)に書き込むことで実現される。もし、HDD 1 1-0が実際に故障した後であれば、HDD 1 1-2でのメディアエラーは修復できず、したがってHDD 1 1-1及び1 1-2からHDD 1 1-0のデータを復元することもできなくなる。なお、メディアエラー修復部26により修復で、HDD 1 1-2内部で代替処理が行われることがある。この場合、復元されたデータが書き込まれる物理位置は、メディアエラーが発生した物理位置とは異なる。但し、代替先の物理位置は、メディアエラーが発生したアドレスでリンク付けされており、当該アドレスで正しくアクセスされる。

10

【0037】

HDD健全性確認部25は、メディアエラー発生箇所をメディアエラー修復部26により修復させると、健全性確認の対象とすべきアドレスを次のデータブロックを指定するように更新してステップS 1 1に戻る。

20

【0038】

以上のようにして、論理ディスク10を構成するHDD 1 1-0~1 1-2のうち、故障が予知されたHDD 1 1-0を除く全てのHDD 1 1-j (HDD 1 1-1及び1 1-2)を対象とする健全性の確認処理が進んだ結果、次に健全性確認の対象とすべきアドレスが、図3(b)に示すAcとなったものとする。このとき、データコピー部27によるHDD 1 1-0からスペアHDD 1 1-3へのデータコピー処理は、アドレス0から上記アドレスAcまで完了しているものとする。つまり、次に健全性確認の対象とすべきアドレスが、スペアHDD 1 1-3へのデータコピー完了領域に到達した、更に詳細に述べるならば、当該データコピー完了領域に相当するHDD 1 1-j (HDD 1 1-1及び1 1-2)の領域に到達したものとする。

30

【0039】

HDD健全性確認部25は、HDD 1 1-j (HDD 1 1-1及び1 1-2)の全領域の健全性を確認(スキャン)し終えない場合でも(ステップS 11)、次に健全性確認の対象とすべきアドレスが上記データコピー完了領域に到達したならば(ステップS 12)、健全性確認処理を終了する。その理由は、既に述べたように、HDD 1 1-3のデータコピー完了領域(アドレス0~Acまでの領域)に相当するHDD 1 1-1または1 1-2の領域に、仮にメディアエラーが発生する箇所が含まれていて、その箇所が修復されなかったとしても、その箇所のデータは、新たな論理ディスク10を構成する、HDD 1 1-3を含む他のHDDのデータから復元可能であるためである。なお、健全性確認処理をHDD 1 1-j (HDD 1 1-1及び1 1-2)の全領域について実行しても構わない。

40

【0040】

上記実施形態では、HDD故障予知部24によりHDD 1 1-0の故障が予知されてから、当該HDD 1 1-0からスペアHDD 1 1-3へのデータコピーが完了するまでに要する時間を短縮できるように、HDD健全性確認部25による健全性確認処理とデータコピー部27によるコピー処理とがほぼ同時に、且つアドレスを更新する方向が相互に逆順となるようにしている。しかし、故障が予知されたHDD 1 1-0のデータを当該HDD 1 1-0が実際に故障する前にスペアHDD 1 1-3にコピーすると共に、当該HDD 1 1-0のデータを利用して残りのHDD 1 1-1及び1 1-2のメディアエラーを修復するという観点からは、この手法に限らない。例えば、HDD健全性確認部25によるHDD 1 1-1及び1 1-2の全領域を対象に健全性確認処理を先に行い、しかる後にデータコピー部27によるコピ

50

ー処理を行うようにしても良い。また、HDD 11-0のデータまたは冗長データを利用して残りのHDD 11-1及び11-2のメディアエラーを修復するという観点だけに着目するならば、データコピー部27によるコピー処理は必ずしも必要ない。この場合には、HDD 11-0をスペアHDD 11-3で代替する際に、論理ディスク復元部23が、HDD 11-0のデータをHDD 11-1及び11-2のデータで復元してHDD 11-3に書き込めばよい。

【0041】

また上記実施形態では、データコピー部27は、故障が予知されたHDD 11-0からデータが読み出せない場合を除いて、当該HDD 11-0のデータをそのままスペアHDD 11-3にコピーしている。しかし、論理ディスク10を構成するHDD 11-0~11-2のうち、残りのHDD 11-1及び11-2のデータまたは冗長データを用いてHDD 11-0のデータを復元して、その復元されたデータをHDD 11-3にコピーするようにしても構わない。

10

【0042】

なお、本発明は、上記実施形態そのままに限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で構成要素を変形して具体化できる。また、上記実施形態に開示されている複数の構成要素の適宜な組み合わせにより種々の発明を形成できる。例えば、実施形態に示される全構成要素から幾つかの構成要素を削除してもよい。

【図面の簡単な説明】

【0043】

【図1】本発明の一実施形態に係るディスクアレイ装置の構成を示すブロック図。

【図2】HDD故障予知部24によってHDD 11-iの故障が予知された場合の処理の手順を示すフローチャート。

【図3】HDD健全性確認部25によるHDD健全性確認処理とデータコピー部27によるデータコピー処理の各々の処理が進行するアドレスの方向と、HDD健全性確認処理が終了するアドレスとを示す図。

【図4】HDD健全性確認部25によるHDD健全性確認処理の詳細な手順を示すフローチャート。

【符号の説明】

【0044】

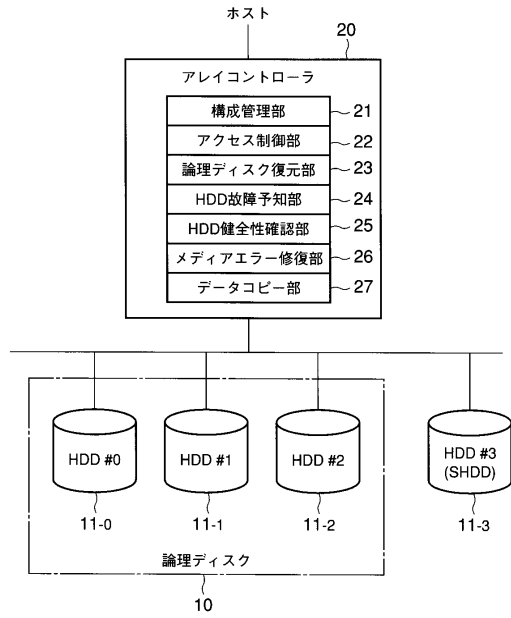
10...論理ディスク、11-0~11-2...HDD(磁気ディスクドライブ)、11-3...スペアのHDD(磁気ディスクドライブ)、20...アレイコントローラ、21...構成管理部、22...アクセス制御部、23...論理ディスク復元部、24...HDD故障予知部、25...HDD健全性確認部、26...メディアエラー修復部、27...データコピー部。

20

30

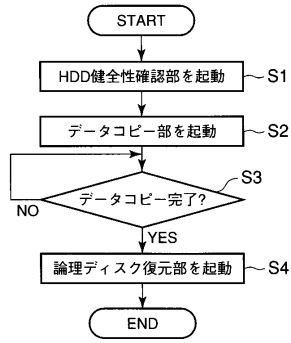
【図1】

図1



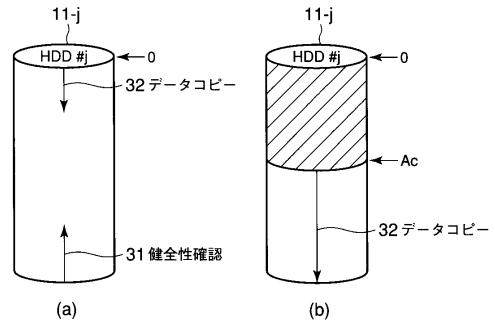
【図2】

図2



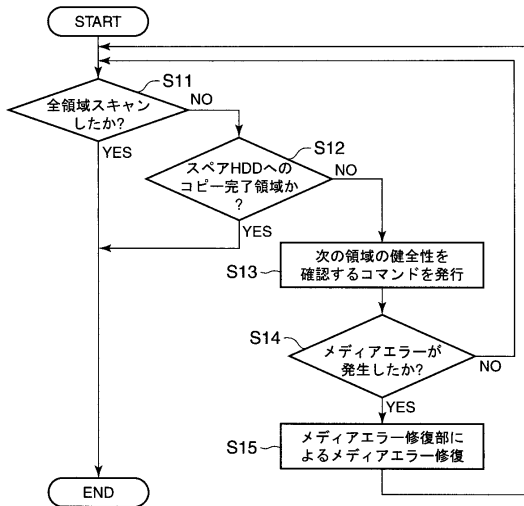
【図3】

図3



【図4】

図4



フロントページの続き

(74)代理人 100075672

弁理士 峰 隆司

(74)代理人 100109830

弁理士 福原 淑弘

(74)代理人 100084618

弁理士 村松 貞男

(74)代理人 100092196

弁理士 橋本 良郎

(72)発明者 友田 正憲

東京都港区芝浦一丁目1番1号 東芝ソリューション株式会社内

審査官 横山 佳弘

(56)参考文献 特開2003-303057(JP,A)

特開2001-100949(JP,A)

特開平09-016340(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F 3/06

G06F 13/10