

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2009-217362

(P2009-217362A)

(43) 公開日 平成21年9月24日(2009.9.24)

(51) Int.Cl.		F I			テーマコード (参考)
G06F 3/06	(2006.01)	G06F 3/06	304R		5B014
G06F 13/10	(2006.01)	G06F 3/06	540		5B065
		G06F 13/10	340A		

審査請求 未請求 請求項の数 6 O L (全 6 頁)

(21) 出願番号 特願2008-58104 (P2008-58104)
 (22) 出願日 平成20年3月7日(2008.3.7)

(71) 出願人 000004237
 日本電気株式会社
 東京都港区芝五丁目7番1号
 (74) 代理人 100084250
 弁理士 丸山 隆夫
 (72) 発明者 小川 和宏
 東京都港区芝五丁目7番1号 日本電気株式会社内
 Fターム(参考) 5B014 EA04 EA05 EB04
 5B065 BA01 CA30 EA31 EK07

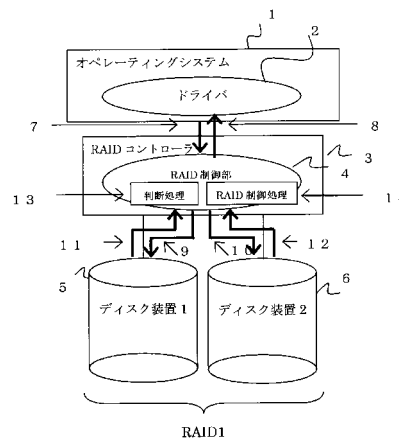
(54) 【発明の名称】 ディスクアレイ装置および制御方法

(57) 【要約】

【課題】 故障の前兆のディスク装置を縮退させることなく、RAIDの冗長性を保ったまま、システムの性能低下を防ぐ制御方法等を提供すること。

【解決手段】 RAID1構成のディスク装置からのコマンド応答時間を検出し、応答時間からディスクの故障の前兆を判断し、故障予測をしたときにプライマリ/セカンダリ制御により冗長性を保持する。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

R A I D 1 構成のディスク装置からのコマンド応答時間を検出する手段と、応答時間からディスクの故障の前兆を判断する手段と、故障予測をしたときにプライマリ/セカンダリ制御により冗長性を保持する手段を備えることを特徴とするディスクアレイ装置。

【請求項 2】

ディスク装置は R A I D 5 構成を備えることを特徴とする請求項 1 記載のディスクアレイ装置。

【請求項 3】

ディスク装置からのレスポンスから故障の前兆の有無を判断する判断処理手段を有し、前兆があった場合は縮退させずに R A I D 1 の状態で冗長性を保ったまま、リード処理をロードバランスからプライマリ/セカンダリの制御に移行する R A I D 制御移行処理手段を有し、リード処理はプライマリのディスク装置からのみ実行することを特徴とする請求項 1 または 2 記載のディスクアレイ装置。

10

【請求項 4】

R A I D 1 構成のディスク装置からのコマンド応答時間を検出するステップと、応答時間からディスクの故障の前兆を判断するステップと、故障予測をしたときにプライマリ/セカンダリ制御により冗長性を保持するステップとを有することを特徴とする制御方法。

【請求項 5】

ディスク装置は R A I D 5 構成を備えることを特徴とする請求項 4 記載の制御方法。

20

【請求項 6】

ディスク装置からのレスポンスから故障の前兆の有無を判断する判断処理ステップを有し、前兆があった場合は縮退させずに R A I D 1 の状態で冗長性を保ったまま、リード処理をロードバランスからプライマリ/セカンダリの制御に移行する R A I D 制御移行処理ステップを有し、リード処理はプライマリのディスク装置からのみ実行することを特徴とする請求項 4 または 5 記載の制御方法。

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、R A I D 1 のシステムにおいて R A I D 1 を構成しているディスク装置のレスポンスから故障の前兆を判断し、前兆があった場合は、縮退させずに R A I D の制御方法をロードバランスからプライマリ/セカンダリ制御に移行することで故障の前兆があるディスクに起因したリード性能の低下を防ぎ、R A I D 1 の状態で冗長性を保つ技術に関する。

30

【背景技術】**【0002】**

R A I D 1 のリード処理は、R A I D を構成する 2 台のディスク装置にロードバランスして I / O を分散させるのが一般的である。もし、リード対象のディスク装置に故障の前兆があった場合、I / O エラーやリトライが発生することがあり、そのためリード処理が完了するまでに時間がかかってしまい、システム性能が低下する問題がある。R A I D はリトライで救えるようなケースを故障と判断しないため、故障と判断して H D D を切り離すまでは、このシステム性能の低下が継続することになる。

40

【0003】

関連技術として、プライマリ・ボリュームがオフラインになると、たとえディスク・システム内にいくつかのセカンダリ・ボリュームが備わっている場合であっても、そのオンライン・トランザクションを中止することが可能な技術が提案されている。その構成は、ディスクの複数グループとして組織化された複数の物理ドライブを包含するマルチ・ディスク・システム。各グループは、複数の論理ボリュームとしてアクセスされる。論理ドライブの 1 つはプライマリ・ボリュームと見なされ、そのほかの 1 つないしは複数の論理ボリュームはセカンダリ・ボリュームと見なされ、これらはともにミラーリング・グループを

50

構成する。メモリ内に収められるデータが、これらの編成を記録している。プライマリ・ボリュームを含むディスク・グループに障害が発生すると、セカンダリ・ボリュームを選択するためにこのデータが参照される（例えば、特許文献1参照）。

【特許文献1】特開2002-182862号公報

【発明の開示】

【発明が解決しようとする課題】

【0004】

上述の問題を回避するためにディスク装置の故障を判断する基準を厳しくして切り離すようにすると、RAIDの縮退が頻繁に発生し、RAIDの冗長性が失われてしまう問題がある。

10

【0005】

第1の課題は、ディスク故障の前兆は故障ではないため、縮退させないことである。

【0006】

第2の課題は、縮退させないため、故障の前兆があるディスク装置に対してI/Oが発行され、システムの性能低下が発生することである。

【0007】

第3の課題は、故障の前兆のディスク装置を縮退させると縮退は頻繁に発生し、RAIDの冗長性が失われてしまうことである。

【0008】

本発明は、以上のような課題を解決するためになされたもので、故障の前兆のディスク装置を縮退させることなく、RAIDの冗長性を保ったまま、システムの性能低下を防ぐディスクアレイ装置および制御方法を提供することを目的とする。

20

【課題を解決するための手段】

【0009】

本発明のディスクアレイ装置は、RAID1構成のディスク装置からのコマンド応答時間を検出する手段と、応答時間からディスクの故障の前兆を判断する手段と、故障予測をしたときにプライマリ/セカンダリ制御により冗長性を保持する手段を備えることを特徴とする。

【0010】

また、本発明の制御方法は、RAID1構成のディスク装置からのコマンド応答時間を検出するステップと、応答時間からディスクの故障の前兆を判断するステップと、故障予測をしたときにプライマリ/セカンダリ制御により冗長性を保持するステップとを有することを特徴とする。

30

【発明の効果】

【0011】

本発明によれば、故障の前兆のディスク装置を縮退させることなく、RAIDの冗長性を保ったまま、システムの性能低下を防ぐディスクアレイ装置および制御方法を提供することが可能となる。

【発明を実施するための最良の形態】

【0012】

以下、本発明の第一の実施の形態について図面を参照して詳細に説明する。図1にRAID1のシステムコンポーネント図として示す本実施の形態におけるRAID1の制御システムは、ディスク装置1(5)とディスク装置2(6)の2台のディスク装置から構成される。

40

【0013】

(1)はオペレーティングシステム、(2)はRAIDコントローラ(3)を制御するドライバ、(3)はRAID1を制御するRAIDコントローラ、(4)はRAIDコントローラ上でRAIDの制御処理を行うRAID制御部、(5)はRAID1を構成するディスク装置1、(6)はRAID1を構成するディスク装置2である。

【0014】

50

(7)はドライバ(2)からRAID1に対して発行されたI/O、(8)はRAID1に発行したI/Oのレスポンス、(9)はディスク装置(5)に発行されたI/O、(10)はディスク装置(6)に発行されたI/O、(11)はディスク装置(5)に発行されたI/Oのレスポンス、(12)はディスク装置(6)に発行されたI/Oのレスポンス、(13)はレスポンス((11)、(12))からディスク故障の前兆を判断する処理、(14)はRAID1の制御方法をロードバランスからプライマリ/セカンダリの制御に移行する処理を示す。

【0015】

RAID1に対するI/Oは、オペレーティングシステム(1)上のドライバ(2)からRAIDコントローラ(3)のRAID制御部(4)に発行(7)される。リード処理のI/Oは、RAID制御部(4)がロードバランスしてディスク装置1(5)かディスク装置2(6)のどちらか((9)または(10))に発行する。そのレスポンス((11)または(12))はRAID制御部((14))からドライバ((12))に戻る((8))。

10

【0016】

以下に図面を参照して本実施の形態の動作の詳細に説明する。

図1において、ディスク装置1(5)に故障の前兆がある場合にディスク装置1(5)にリード処理(9)を発行すると、レスポンス(11)がエラーになったり、レスポンス(11)が返るまでに時間がかかったりすることがある。レスポンス(11)が正常終了でない場合は、RAID制御部(4)のエラー処理としてリトライ(再度(9)を発行)などを実行する。そのため、ドライバ(2)にI/Oの完了が返る(8)までに時間がかかり、システム性能が低下する。RAID制御部(4)は、エラー処理が実行された場合でもエラー処理の結果が正常であれば、ディスク装置(5)を縮退させないため、再度、ディスク装置1(5)にリード処理のI/Oが発行(9)された場合、同様にシステム性能の低下が発生する。

20

【0017】

関連技術では故障の前兆があるディスク装置1(5)を縮退させることになるが、縮退させた場合、RAIDの冗長性が失われてしまう。また故障の判断基準が厳しかった場合、頻繁に縮退が発生する可能性がある。

【0018】

本実施の形態では、RAID制御部(4)に判断処理(13)とRAID制御移行処理(14)を組み込み、判断処理(13)でレスポンス((11)または(12))の状況から故障の前兆があるか判断する。前兆がある場合は縮退させずに、RAID制御移行処理(14)によりロードバランスを中止して、ディスク装置1(5)に故障の前兆がある場合は、ディスク装置1(5)をセカンダリ、残りのもう1台のディスク装置2(6)をプライマリとし、ディスク装置2(6)に故障の前兆がある場合は、ディスク装置2(6)をセカンダリ、残りのもう1台のディスク装置1(5)をプライマリとする、プライマリ/セカンダリ制御に移行する。プライマリ/セカンダリ制御移行後は、リード処理はプライマリのディスク装置のみに発行する。故障の前兆があるディスク装置はセカンダリとし、リード処理は発行しないため、システム性能の低下は発生しない。

30

40

【0019】

図2は、本発明の処理の流れを示した図である。

ドライバはリード処理のI/OをRAIDコントローラに発行(1)する。I/OはRAIDコントローラ上のRAID制御部が受け取り、リード処理のI/Oを発行するディスク装置をロードバランスで決定(2)し、発行(3)する。ディスク装置で処理されたI/OはRAID制御部に返る(4)。RAID制御部では返ってきたレスポンスが正常か異常か判断(5)する。

【0020】

判断(5)の結果が異常(6)の場合は、RAID制御移行処理でロードバランス制御からプライマリ/セカンダリ制御への変更、異常レスポンスが返ってきたディスク装置を

50

セカンダリ、残りのディスク装置をプライマリに設定(7)する。プライマリ/セカンダリ制御にRAID制御が移行したことをRAID制御部で保持(8)する。その後、リード処理のI/O結果をドライバに返し(9)、ドライバでI/Oを受けて取り(11)、リード処理は終了する。

【0021】

判断(5)の結果が正常(10)の場合は、RAID制御部からリード処理のI/O結果をドライバに返し(9)、ドライバでI/Oを受け取り(11)、リード処理は終了する。

【0022】

プライマリ/セカンダリ制御に移行後は、ドライバから発行(12)されたリード処理のI/Oは、RAID制御部が受け取り、リード処理のI/Oをプライマリに設定されているディスク装置(13)に発行(14)する。プライマリのディスク装置で処理されたI/OはRAID制御部に返る(15)。RAID制御部はリード処理のI/O結果をドライバに返し(16)、ドライバでI/Oを受け取り(17)、リード処理は終了する。

【0023】

以上説明したように、本実施の形態においては、以下に記載するような効果を奏する。

第1の効果は、ディスク装置の故障の前兆を判断し、前兆がある場合でも縮退させずにロードバランスからプライマリ/セカンダリの制御に移行することで、RAID1の冗長性を保つことである。

【0024】

第2の効果は、ロードバランスからプライマリ/セカンダリの制御に移行しているため、故障の前兆のあるディスク装置がある場合でもシステムの性能低下が発生しないことである。

【0025】

本発明の他の実施の形態として、他のRAIDレベルでも類似の効果をえられる。

例えば、RAID5はデータとパリティに分散してデータを記憶しており、1台のディスク装置に故障の前兆があった場合、故障のディスク装置にはリードを発行しないようにして、残りのディスク装置からデータ、パリティをリードしてデータを生成することで、同様の効果をえられる。ただし、データの生成処理の時間が余分に必要になるため、僅かではあるが性能は低下する。

【0026】

なお、上述する各実施の形態は、本発明の好適な実施の形態であり、本発明の要旨を逸脱しない範囲内において種々変更実施が可能である。

【図面の簡単な説明】

【0027】

【図1】本発明の実施の形態に係るシステム構成を示す図である。

【図2】本発明の実施の形態に係る処理動作を示すフローチャートである。

【符号の説明】

【0028】

- 1 オペレーティングシステム
- 2 ドライバ
- 3 RAIDコントローラ
- 4 RAID制御部
- 5 ディスク装置1
- 6 ディスク装置2

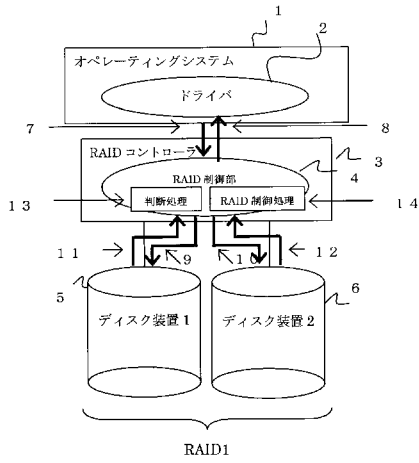
10

20

30

40

【 図 1 】



【 図 2 】

