(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4606455号
(P4606455)
(45)発行日　平成23年1月5日(2011.1.5)　(24)登録日　平成22年10月15日(2010.10.15)

(51)Int.Cl.　　F I
G06F 13/10 (2006.01)　　G06F 13/10 340A
G06F 3/06 (2006.01)　　G06F 3/06 304R

請求項の数 7　(全 29 頁)

(21)出願番号　特願2007-328396(P2007-328396)
(22)出願日　平成19年12月20日(2007.12.20)
(65)公開番号　特開2009-151519(P2009-151519A)
(43)公開日　平成21年7月9日(2009.7.9)
審査請求日　平成21年3月19日(2009.3.19)

(73)特許権者　000005223
富士通株式会社
神奈川県川崎市中原区上小田中４丁目１番１号
(74)代理人　100092152
弁理士　服部　毅巖
(72)発明者　田村　雅寿
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
(72)発明者　野口　泰生
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
(72)発明者　荻原　一隆
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
最終頁に続く

(54)【発明の名称】ストレージ管理装置、ストレージ管理プログラムおよびストレージシステム

(57)【特許請求の範囲】
【請求項１】
データをネットワークで接続された複数のストレージ装置に分散して記憶するストレージシステムにおけるストレージ装置を管理するストレージ管理装置において、
前記ストレージ装置の記憶領域が正常に稼動していることを確認するパトロール処理を実行するパトロール処理実行手段と、
前記ストレージ装置における読み出し処理と書き込み処理との処理性能を示すパトロール性能を測定するパトロール性能測定手段と、
前記パトロール性能測定手段によって測定された前記パトロール性能に基づいてパトロール流量設定値を決定し、決定した当該パトロール流量設定値を超えないように、前記パトロール処理実行手段によって実行されている前記パトロール処理の速さであるパトロール流量を規制するパトロール流量規制手段と、
を有することを特徴とするストレージ管理装置。
【請求項２】
前記ストレージ装置と、当該ストレージ装置に記憶されている前記データと同一内容の冗長データを記憶している二重化ストレージ装置とを対応付ける管理情報を記憶する管理情報記憶手段を有し、
前記パトロール処理実行手段は、さらに、前記パトロール処理において、前記管理情報記憶手段に記憶されている前記管理情報を参照して、前記ストレージ装置に記憶されている前記データと、前記二重化ストレージ装置に記憶されている前記データと同一内容の前

記冗長データとを読み出し、読み出した前記データと前記冗長データとを比較することによって、比較した前記データと前記冗長データとが一致することを確認することを特徴とする請求項１記載のストレージ管理装置。

【請求項３】

前記パトロール性能測定手段は、前記ストレージ装置の記憶領域が複数に区分された領域である測定単位領域ごとに、前記パトロール性能を測定し、

前記パトロール処理実行手段は、前記パトロール処理を、前記ストレージ装置の記憶領域が前記測定単位領域からさらに複数に区分された領域である実行単位領域ごとに実行し、

前記パトロール流量規制手段は、前記実行単位領域ごとに前記パトロール処理の前記パトロール流量を規制することを特徴とする請求項１記載のストレージ管理装置。

【請求項４】

現在の時刻を示す時刻情報を取得する時刻情報取得手段を有し、

前記パトロール性能測定手段は、前記時刻情報取得手段によって取得された前記時刻情報が示す時刻が属する時間帯ごとに、測定した前記パトロール性能を複数種類保持し、

前記パトロール流量規制手段は、前記時刻情報取得手段によって取得された前記時刻情報が示す前記現在の時刻に基づいて、前記パトロール性能測定手段が保持する前記パトロール性能から前記現在の時刻が属する時間帯に対応する前記パトロール性能を選択し、選択した当該パトロール性能に基づいて前記パトロール流量設定値を決定し、決定した当該パトロール流量設定値を超えないように前記パトロール流量を規制することを特徴とする請求項１記載のストレージ管理装置。

【請求項５】

前記ストレージ装置と、当該ストレージ装置に記憶されている前記データと同一内容の冗長データを記憶している二重化ストレージ装置とを対応付ける管理情報を記憶する管理情報記憶手段を有し、

前記パトロール処理実行手段は、前記パトロール処理として、前記管理情報記憶手段に記憶されている前記管理情報を参照して、前記二重化ストレージ装置に記憶されている前記データと同一内容の前記冗長データとを読み出し、読み出した前記冗長データを前記ストレージ装置内の前記データの記憶領域に対して上書きで書き込む処理を実行することを特徴とする請求項１記載のストレージ管理装置。

【請求項６】

コンピュータに、データをネットワークで接続された複数のストレージ装置に分散して記憶するストレージシステムにおけるストレージ装置の管理処理を実行させるストレージ管理プログラムにおいて、

前記コンピュータを、

前記ストレージ装置の記憶領域が正常に稼動していることを確認するパトロール処理を実行するパトロール処理実行手段、

前記ストレージ装置における読み出し処理と書き込み処理との処理性能を示すパトロール性能を測定するパトロール性能測定手段、

前記パトロール性能測定手段によって測定された前記パトロール性能に基づいてパトロール流量設定値を決定し、決定した当該パトロール流量設定値を超えないように、前記パトロール処理実行手段によって実行されている前記パトロール処理の速さであるパトロール流量を規制するパトロール流量規制手段、

として機能させることを特徴とするストレージ管理プログラム。

【請求項７】

データをネットワークで接続された複数のストレージ装置に分散して記憶し、前記ストレージ装置を管理するストレージ管理装置を有するストレージシステムにおいて、

前記ストレージ装置と、当該ストレージ装置に記憶されている前記データと同一内容の冗長データを記憶している二重化ストレージ装置とを対応付ける管理情報を記憶する管理情報記憶手段を有し、

前記ストレージ管理装置は、
前記ストレージ装置の記憶領域が正常に稼動していることを確認するパトロール処理を実行するパトロール処理実行手段と、
前記ストレージ装置における読み出し処理と書き込み処理との処理性能を示すパトロール性能を測定するパトロール性能測定手段と、
前記パトロール性能測定手段によって測定された前記パトロール性能に基づいてパトロール流量設定値を決定し、決定した当該パトロール流量設定値を超えないように、前記パトロール処理実行手段によって実行されている前記パトロール処理の速さであるパトロール流量を規制するパトロール流量規制手段と、
を有し、
前記パトロール処理実行手段は、さらに、前記パトロール処理において、前記管理情報記憶手段に記憶されている前記管理情報を参照して、前記ストレージ装置に記憶されている前記データと、前記二重化ストレージ装置に記憶されている前記データと同一内容の前記冗長データとを読み出し、読み出した前記データと前記冗長データとを比較することによって、比較した前記データと前記冗長データとが一致することを確認することを特徴とするストレージシステム。

【発明の詳細な説明】
【技術分野】
【０００１】
本発明はストレージ管理装置、ストレージ管理プログラム、およびストレージシステムに関し、特にデータを複数のストレージノードで分散管理するストレージ管理装置、ストレージ管理プログラム、およびストレージシステムに関する。
【背景技術】
【０００２】
現在、コンピュータを用いたデータ処理が広く行われるようになり、データを蓄積・利用するためのストレージ技術が一層重要となっている。この点、データアクセスの高速化およびデータ保管に対する高信頼化を実現するストレージ技術として、従来はＲＡＩＤ（Redundant Arrays of Independent Disks）が一般的に利用されてきた。ＲＡＩＤでは、データを必要に応じて分割・複製して、複数のストレージ装置に分散して配置する。これにより、複数のストレージ装置間で負荷が分散されることによる高速化、データが冗長化されることによる高信頼化が実現されている。
【０００３】
そして近年では、さらに高速化・高信頼化を図るため、ＲＡＩＤの考え方を応用した分散ストレージシステムが構築されるようになっている。分散ストレージシステムは、複数のストレージノードと、ストレージノード間を接続するネットワークを備えている。各ストレージノードは、内部に備えた、または外部に設けられて対応付けられたストレージ装置を管理しているとともに、ネットワーク通信機能を備えている。分散ストレージシステムは、この複数のストレージノードにデータを分散して記憶することで、システム全体として一層の高速化・高信頼化を実現している。
【０００４】
このようなストレージシステムの有するストレージ装置が正常であるか否かを診断するためには、一般に、稼働しているかどうかの診断（いわゆる生存確認）が行われる。この生存確認による診断は、処理が比較的短時間で可能であるとともに、処理の負荷が小さいため、処理によるシステムの通常の業務のアクセスに対する影響も比較的小さいというメリットがある。
【０００５】
このようなストレージシステムを始めとする、コンピュータシステムの性能の監視に関して、ストレージネットワークの構成要素から収集した性能情報に基づいて、以降の情報収集の対象範囲や程度を自動調整するシステムが知られている（例えば、特許文献１参照

）。
【０００６】
また、ネットワークノードの負荷に応じて監視トラフィックの発生間隔を自動調整するネットワーク管理装置が知られている（例えば、特許文献２参照）。
また、管理対象である計算機の監視項目ごとの稼働性能値に基づいて、採取すべき監視項目を特定する管理システムが知られている（例えば、特許文献３参照）。
【特許文献１】特開２００５－１５７９３３号公報
【特許文献２】特開平０９－２７０７９４号公報
【特許文献３】特開２００４－２０６４９５号公報
【発明の開示】
【発明が解決しようとする課題】
【０００７】
しかし、上記の稼働しているかどうかの診断を行ったのみでは、ストレージ装置自体が稼動していることは確認できるが、ストレージ装置内部の特定の領域のみに発生した障害を検出することは不可能である。
【０００８】
特に、例えばストレージ装置がＲＡＩＤ５によって構成されている場合のように、ストレージ装置間で冗長構成がとられている場合には、一つの領域の障害が検出されれば、これを復旧することが可能である。しかし、冗長構成によって対応付けられている二つの領域について、障害が同時に発生した場合には、障害によって失われたデータを復旧することが不可能である。これにより、ストレージ装置が稼働しているかどうかの診断だけでなく、ストレージ装置の全領域にわたって、障害が発生しているかどうかについても診断することが重要である。
【０００９】
しかし、ストレージ装置が有する記憶領域の全領域について診断することは、ストレージシステムに与える負荷が大きい上に、処理の終了まで長時間を要する。このため、ストレージ装置の記憶領域の診断は、通常の業務によるアクセスへの影響が無視できなくなるという問題点がある。
【００１０】
また、分散ストレージシステムにおいては、多種多様なストレージ装置が存在する場合もある。このような場合には、同じ内容の診断処理を行っても、ストレージ装置の種類によって、その所要時間や通常の業務によるアクセスへの影響が異なる場合がある。このため、ストレージ装置の全領域の診断を行った場合の通常の業務に対する影響の評価が困難であるという問題点がある。
【００１１】
また、分散ストレージシステムにおいては、ストレージノード間で冗長構成がとられる場合もある。データが正常かどうかを診断するためには、冗長構成をとっている各ストレージノードのデータが正しく冗長構成を保持できているかどうかの診断を行う必要がある。ここで、この分散ストレージシステムは、多種多様な構成のストレージノードが混在することによってシステムが構築されている場合もある。この場合、ストレージ装置と同様に、同じ内容の診断処理を行っても、各ストレージノードの構成によって、その所要時間や通常の業務によるアクセスへの影響が異なる場合がある。このため、ストレージ装置の全領域の診断を行った場合の通常の業務に対する影響の評価が困難であるという問題点がある。
【００１２】
しかし、特許文献１～３に記載されている技術では、システムを構成するコンピュータに対するアクセス頻度などの、システムの様々な性能情報を収集することが可能であるが、それ自体では、ストレージ装置の各領域の異常を診断することはできず、また、直接的に診断の負荷を軽減することもできない。
【００１３】

本発明はこのような点に鑑みてなされたものであり、ストレージ装置へのアクセスに対する、パトロール処理の実行に基づく負荷による影響を抑えつつ、ストレージシステムの信頼性を高めることができるストレージ管理装置、ストレージ管理プログラムおよびストレージシステムを提供することを目的とする。

【課題を解決するための手段】

【００１４】

この、データをネットワークで接続された複数のストレージ装置に分散して記憶するストレージシステムにおけるストレージ装置を管理するストレージ管理装置は、前記ストレージ装置の記憶領域が正常に稼動していることを確認するパトロール処理を実行するパトロール処理実行手段と、前記パトロール処理実行手段によって実行されている前記パトロール処理の速さであるパトロール流量を規制するパトロール流量規制手段と、を有することを要件とする。

【００１５】

このようなストレージ管理装置によれば、パトロール処理実行手段により、ストレージ装置の記憶領域が正常に稼動していることを確認するパトロール処理が実行される。また、パトロール流量規制手段により、パトロール処理実行手段によって実行されているパトロール処理の速さであるパトロール流量が規制される。

【００１６】

また、コンピュータに、データをネットワークで接続された複数のストレージ装置に分散して記憶するストレージシステムにおけるストレージ装置の管理処理を実行させるストレージ管理プログラムは、前記コンピュータを、前記ストレージ装置の記憶領域が正常に稼動していることを確認するパトロール処理を実行するパトロール処理実行手段、前記パトロール処理実行手段によって実行されている前記パトロール処理の速さであるパトロール流量を規制するパトロール流量規制手段、として機能させることを要件とする。

【００１７】

このようなストレージ管理プログラムを実行するコンピュータによれば、パトロール処理実行手段により、ストレージ装置の記憶領域が正常に稼動していることを確認するパトロール処理が実行される。また、パトロール流量規制手段により、パトロール処理実行手段によって実行されているパトロール処理の速さであるパトロール流量が規制される。

【００１８】

また、データをネットワークで接続された複数のストレージ装置に分散して記憶し、前記ストレージ装置を管理するストレージ管理装置を有するストレージシステムは、前記ストレージ装置と、当該ストレージ装置に記憶されている前記データと同一内容の冗長データを記憶している二重化ストレージ装置とを対応付ける管理情報を記憶する管理情報記憶手段を有し、前記ストレージ管理装置は、前記ストレージ装置の記憶領域が正常に稼動していることを確認するパトロール処理を実行するパトロール処理実行手段と、前記パトロール処理実行手段によって実行されている前記パトロール処理の速さであるパトロール流量を規制するパトロール流量規制手段と、を有し、前記パトロール処理実行手段は、さらに、前記パトロール処理において、前記管理情報記憶手段に記憶されている前記管理情報を参照して、前記ストレージ装置に記憶されている前記データと、前記二重化ストレージ装置に記憶されている前記データと同一内容の前記冗長データとを読み出し、読み出した前記データと前記冗長データとを比較することによって、比較した前記データと前記冗長データとが一致することを確認することを要件とする。

【００１９】

このようなストレージシステムによれば、管理情報記憶手段により、当該ストレージ装置に記憶されているデータと同一内容の冗長データを記憶している二重化ストレージ装置とを対応付ける管理情報が記憶されている。そして、パトロール処理実行手段により、ストレージ装置の記憶領域が正常に稼動していることを確認するパトロール処理が実行される。また、パトロール流量規制手段により、パトロール処理実行手段によって実行されているパトロール処理の速さであるパトロール流量が規制される。さらに、パトロール処理

実行手段により、パトロール処理において、管理情報記憶手段に記憶されている管理情報が参照され、ストレージ装置に記憶されているデータと、二重化ストレージ装置に記憶されているデータと同一内容の冗長データとが読み出され、読み出されたデータと冗長データとを比較することによって、比較されたデータと冗長データとが一致することが確認される。

【発明の効果】

【００２０】

開示のストレージ管理装置、ストレージ管理プログラムおよびストレージシステムによれば、パトロール処理の実行に基づく負荷による、ストレージ装置へのアクセスに対する影響を抑えつつ、ストレージシステムの信頼性を高めることができる。

【発明を実施するための最良の形態】

【００２１】

以下、本発明の実施の形態について、図面を参照して説明する。

本実施の形態のストレージ管理装置１は、多種多様なストレージ装置を管理する多種多様なストレージノードがネットワークでつながれているマルチノードストレージシステムにおいて、ストレージ装置の各記憶領域へのデータの読み書きが正常に行えるかを診断するパトロール処理を、自律的に調整して実行することで、アクセスへの影響を抑えつつ、ストレージシステムの信頼性を向上させるものである。

【００２２】

図１は、本実施の形態の概要を示す図である。図１に示すストレージシステムは、データをネットワークで接続された複数のストレージ装置に分散して記憶するストレージシステムである。このストレージシステムは、ストレージ管理装置１、ストレージ装置２から構成される。

【００２３】

ストレージ管理装置１は、ストレージ装置２を管理する装置である。ストレージ管理装置１は、ストレージ装置２を管理するために、パトロール処理実行手段１ａ、およびパトロール流量規制手段１ｂを有している。

【００２４】

パトロール処理実行手段１ａは、ストレージ装置２の記憶領域２ａが正常に稼動していることを確認するパトロール処理を実行する。ここで、記憶領域２ａは、ストレージ装置２においてデータを記憶する領域である。このパトロール処理は、具体的には、パトロール処理実行手段１ａが、ストレージ装置２の記憶領域２ａに記憶されているデータを一旦読み出した後、この読み出したデータを、そのデータが記憶されていた領域に対して書き込み、これを記憶領域２ａを区分した領域ごとに順次繰り返すことで行われる。

【００２５】

パトロール流量規制手段１ｂは、パトロール処理実行手段１ａによって実行されているパトロール処理の速さであるパトロール流量を規制する。このパトロール流量の規制は、具体的には、例えば、パトロール流量規制手段１ｂが、パトロール処理実行手段１ａによって実行されるパトロール処理のパトロール流量を監視し、パトロール流量が速過ぎる場合には、パトロール処理実行手段１ａによるパトロール処理の実行を待機させることによって行われる。

【００２６】

このようなストレージ管理装置１によれば、パトロール処理実行手段１ａにより、ストレージ装置２の記憶領域２ａが正常に稼動していることが確認される。パトロール流量規制手段１ｂにより、パトロール処理実行手段１ａで実行されるパトロール処理の速さが規制される。

【００２７】

これによって、パトロール処理の実行に基づく負荷による、ストレージ装置へのアクセスに対する影響を抑えつつ、ストレージシステムの信頼性を高めることができる。

以下、本実施の形態について図面を参照して詳細に説明する。

【００２８】
図２は、本実施の形態のシステム構成を示す図である。図２に示すストレージシステムは、データをネットワークで接続された複数のストレージ装置で分散して記憶する。本実施の形態に係るストレージシステムでは、ストレージノード１００，２００，３００，４００、コントロールノード５００、アクセスノード６００および管理ノード３０がネットワーク１０を介して相互に接続されている。また、端末装置２１，２２，２３が、ネットワーク２０を介してアクセスノード６００に接続されている。
【００２９】
ストレージノード１００，２００，３００，４００には、それぞれストレージ装置１１０，２１０，３１０，４１０が接続されている。ストレージノード１００，２００，３００，４００は、接続されたストレージ装置１１０，２１０，３１０，４１０に格納されたデータを管理し、管理しているデータをネットワーク１０経由でアクセスノード６００に提供する。また、ストレージノード１００，２００，３００，４００は、データに冗長性をもたせて管理している。すなわち、同一内容のデータが、少なくとも２つのストレージノードで管理されている。
【００３０】
ストレージ装置１１０には、ハードディスク装置（ＨＤＤ）１１１，１１２，１１３，１１４が実装されている。ストレージ装置２１０には、ＨＤＤ２１１，２１２，２１３，２１４が実装されている。ストレージ装置３１０には、ＨＤＤ３１１，３１２，３１３，３１４が実装されている。ストレージ装置４１０には、ＨＤＤ４１１，４１２，４１３，４１４が実装されている。ストレージ装置１１０，２１０，３１０，４１０は、内蔵する複数のＨＤＤを用いたＲＡＩＤシステムである。本実施の形態では、ストレージ装置１１０，２１０，３１０，４１０は、ＲＡＩＤ５のディスク管理サービスを提供する。
【００３１】
なお、ストレージ装置１１０，２１０，３１０，４１０は、ＲＡＩＤ５に限らず、他のＲＡＩＤを用いて構成してもよく、ＲＡＩＤ以外の技術を用いてディスクアレイを構成してもよい。さらに、ストレージ装置１１０，２１０，３１０，４１０は、ディスクアレイを用いずにハードディスク単体で構成してもよく、その他の記憶装置を用いてもよい。
【００３２】
ストレージノード１００，２００，３００，４００は、接続されたストレージ装置１１０，２１０，３１０，４１０に格納されたデータを管理し、管理しているデータをネットワーク１０経由で端末装置２１，２２，２３に提供する。また、ストレージノード１００，２００，３００，４００は、冗長性を有するデータを管理している。すなわち、同一のデータが、少なくとも２つのストレージノードで管理されている。
【００３３】
さらに、ストレージノード１００，２００，３００，４００は、パトロール処理において、二重化されたデータの整合性をチェックする。なお、ストレージノード１００，２００，３００，４００は個々の判断に基づいてデータの整合性をチェックしてもよいし、外部からの指示によりデータの整合性をチェックしてもよい。本実施の形態では、各ストレージノード１００，２００，３００，４００が、自発的に二重化されたデータの整合性をチェックする。
【００３４】
パトロール処理における二重化されたデータの整合性のチェックでは、二重化されたそれぞれのデータを保持するストレージノード同士が互いに通信し合い、冗長性のあるデータの整合性がチェックされる。その際、二重化されたデータを保持する一方のストレージノードで管理されているデータで不具合が検出されれば、他方のストレージノードの対応するデータを用いてデータの復旧が行われる。
【００３５】
コントロールノード５００は、ストレージノード１００，２００，３００，４００を管理する。具体的には、コントロールノード５００は、データの配置状況を示す論理ボリュ

ームを保持している。コントロールノード５００は、ストレージノード１００，２００，３００，４００からデータの管理に関する情報を取得し、必要に応じて論理ボリュームを更新する。また、コントロールノード５００は、論理ボリュームが更新されると、その影響を受けるストレージノードに対して更新内容を通知する。論理ボリュームに関しては、後述する。

【００３６】

アクセスノード６００は、端末装置２１，２２，２３に対して、ストレージノード１００，２００，３００，４００が管理するデータを利用した情報処理のサービスを提供する。すなわち、アクセスノード６００は、端末装置２１，２２，２３からの要求に応答して所定のプログラムを実行し、必要に応じてストレージノード１００，２００，３００，４００にアクセスする。ここで、アクセスノード６００は、コントロールノード５００から論理ボリュームを取得し、取得した論理ボリュームに基づいてアクセスすべきストレージノードを特定する。

【００３７】

管理ノード３０は、分散ストレージシステムの管理者が操作する端末装置である。分散ストレージシステムの管理者は、管理ノード３０を操作して、ストレージノード１００，２００，３００，４００、コントロールノード５００およびアクセスノード６００にアクセスし、運用に必要な各種設定を行うことができる。

【００３８】

端末装置２１，２２，２３は、ストレージシステムのユーザが操作するコンピュータである。ストレージシステムのユーザは、端末装置２１，２２，２３を操作して、ストレージノード１００にアクセスする。これに基づいて、ストレージノード１００は、ユーザの要求に応じて、ストレージ装置１１０に記憶されているデータの読み書きを行う。

【００３９】

次に、ストレージノード１００，２００，３００，４００、コントロールノード５００、アクセスノード６００、端末装置２１，２２，２３および管理ノード３０のハードウェア構成について説明する。

【００４０】

図３は、ストレージノードのハードウェア構成例を示す図である。ストレージノード１００は、ＣＰＵ（Central Processing Unit）１０１によって装置全体が制御されている。ＣＰＵ１０１には、バス１０７を介してＲＡＭ（Random Access Memory）１０２、ＨＤＤインタフェース１０３、グラフィック処理装置１０４、入力インタフェース１０５、および通信インタフェース１０６が接続されている。

【００４１】

ＲＡＭ１０２には、ＣＰＵ１０１に実行させるＯＳ（Operating System）のプログラムやアプリケーションプログラムの少なくとも一部が一時的に格納される。また、ＲＡＭ１０２には、ＣＰＵ１０１による処理に必要な各種データが格納される。

【００４２】

ＨＤＤインタフェース１０３には、ストレージ装置１１０が接続されている。ＨＤＤインタフェース１０３は、ストレージ装置１１０に内蔵されたＲＡＩＤコントローラ１１５と通信し、ストレージ装置１１０に対するデータの入出力を行う。ストレージ装置１１０内のＲＡＩＤコントローラ１１５は、ＲＡＩＤ０～５の機能を有し、複数のＨＤＤ１１１～１１４をまとめて１台のハードディスクとして管理する。

【００４３】

グラフィック処理装置１０４には、モニタ１１が接続されている。グラフィック処理装置１０４は、ＣＰＵ１０１からの命令に従って、画像をモニタ１１の画面に表示させる。入力インタフェース１０５には、キーボード１２とマウス１３とが接続されている。入力インタフェース１０５は、キーボード１２やマウス１３から送られてくる信号を、バス１０７を介してＣＰＵ１０１に送信する。

【００４４】

通信インタフェース１０６は、ネットワーク１０に接続されている。通信インタフェース１０６は、ネットワーク１０を介して、他のコンピュータとの間でデータの送受信を行う。

【００４５】

以上のようなハードウェア構成によって、本実施の形態の処理機能を実現することができる。なお、図３には、ストレージノード１００およびストレージ装置１１０の構成のみを示したが、他のストレージノード２００，３００，４００や他のストレージ装置２１０，３１０，４１０も同様のハードウェア構成で実現できる。

【００４６】

さらに、コントロールノード５００、アクセスノード６００、および端末装置２１，２２，２３も、ストレージノード１００とストレージ装置１１０との組合せと同様のハードウェア構成で実現できる。ただし、コントロールノード５００、アクセスノード６００、および端末装置２１，２２，２３については、ストレージ装置１１０のようなＲＡＩＤシステムではなく、単体のＨＤＤがＨＤＤコントローラに接続されていてもよい。

【００４７】

図２に示すように、複数のストレージノード１００，２００，３００，４００がネットワーク１０に接続され、それぞれのストレージノード１００，２００，３００，４００は他のストレージノードとの間で通信を行う。この分散ストレージシステムは、端末装置２１，２２，２３に対して、仮想的なボリューム（以下、論理ボリュームと呼ぶ）として機能する。

【００４８】

図４は、論理ボリュームのデータ構造を示す図である。論理ボリューム７００には、「ＬＶＯＬ－Ａ」という識別子（論理ボリューム識別子）が付与されている。また、ネットワーク経由で接続された４台のストレージ装置１１０，２１０，３１０，４１０には、それぞれのストレージ装置を管理するストレージノードの識別のためにそれぞれ「ＳＮ－Ａ」、「ＳＮ－Ｂ」、「ＳＮ－Ｃ」、「ＳＮ－Ｄ」というノード識別子が付与されている。

【００４９】

各ストレージノード１００，２００，３００，４００が有するストレージ装置１１０，２１０，３１０，４１０それぞれにおいてＲＡＩＤ５の論理ディスクが構成されている。この論理ディスクは、５つのスライスに分割され、個々のストレージノード内で管理されている。

【００５０】

図４の例では、ストレージ装置１１０内の記憶領域は、５つのスライス１２１～１２５に分けられている。ストレージ装置２１０内の記憶領域は、５つのスライス２２１～２２５に分けられている。ストレージ装置３１０内の記憶領域は、５つのスライス３２１～３２５に分けられている。ストレージ装置４１０内の記憶領域は、５つのスライス４２１～４２５に分けられている。

【００５１】

なお、論理ボリューム７００は、セグメント７１０，７２０，７３０，７４０という単位で構成される。セグメント７１０，７２０，７３０，７４０の記憶容量は、ストレージ装置１１０，２１０，３１０，４１０における管理単位であるスライスの記憶容量と同じである。例えば、スライスの記憶容量が１ギガバイトとするとセグメントの記憶容量も１ギガバイトである。論理ボリューム７００の記憶容量はセグメント１つ当りの記憶容量の整数倍である。セグメントの記憶容量が１ギガバイトならば、論理ボリューム７００の記憶容量は４ギガバイトといったものになる。

【００５２】

セグメント７１０，７２０，７３０，７４０は、それぞれプライマリスライス７１１，７２１，７３１，７４１とセカンダリスライス７１２，７２２，７３２，７４２との組から構成される。同じセグメントに属するスライスは別々のストレージノードに属する。個々のスライスを管理する領域には論理ボリューム識別子やセグメント情報や同じセグメン

トを構成するスライス情報の他にフラグがあり、そのフラグにはプライマリあるいはセカンダリなどを表す値が格納される。

【００５３】

図４の例では、スライスの識別子を、「Ｐ」または「Ｓ」のアルファベットと数字との組合せで示している。「Ｐ」はプライマリスライスであることを示している。「Ｓ」はセカンダリスライスであることを示している。アルファベットに続く数字は、何番目のセグメントに属するのかを表している。例えば、１番目のセグメント７１０のプライマリスライスが「Ｐ１」で示され、セカンダリスライスが「Ｓ１」で示される。

【００５４】

このような構造の論理ボリューム７００の各プライマリスライスおよびセカンダリスライスが、ストレージ装置１１０，２１０，３１０，４１０内のいずれかのスライスに対応付けられる。例えば、セグメント７１０のプライマリスライス７１１は、ストレージ装置４１０のスライス４２４に対応付けられ、セカンダリスライス７１２は、ストレージ装置２１０のスライス２２２に対応付けられている。

【００５５】

そして、各ストレージ装置１１０，２１０，３１０，４１０では、自己のスライスに対応するプライマリスライスまたはセカンダリスライスのデータを記憶する。

次に、ストレージノード１００，２００，３００，４００のモジュール構成について説明する。

【００５６】

図５は、ストレージノードの機能を示すブロック図である。なお、図５ではストレージノード１００のモジュール構成を示しているが、他のストレージノード２００，３００，４００もストレージノード１００と同様のモジュール構成によって実現できる。

【００５７】

ストレージノード１００は、ストレージ装置１１０を管理する装置である。ストレージノード１００は、ストレージ装置１１０を管理するために、パトロール処理実行部１３１、パトロール流量規制部１３２、管理情報記憶部１３３、パトロール性能測定部１３４、および時刻情報取得部１３５を有している。同様に、ストレージノード２００は、ストレージ装置２１０を管理する装置である。ストレージノード２００は、ストレージ装置２１０を管理するために、ストレージノード１００と同様の機能を有している。

【００５８】

ここで、ストレージ装置１１０は、ストレージ装置２１０と二重化されているものとする。すなわち、ストレージ装置１１０に記憶されているデータには、ストレージ装置２１０に記憶されている同一の内容の冗長データが存在するものとする。このようにして、ストレージ装置１１０のデータはストレージ装置２１０の冗長データによりバックアップされている。

【００５９】

パトロール処理実行部１３１は、ストレージ装置１１０の記憶領域１１０ａが正常に稼動していることを確認するパトロール処理を実行する。具体的には、パトロール処理実行部１３１は、パトロール処理として、ストレージ装置１１０の記憶領域１１０ａに記憶されているデータを読み出し、データが正常か否かを判断する処理を実行する。ここで、記憶領域１１０ａは、ストレージ装置１１０においてデータを記憶する領域である。

【００６０】

パトロール処理実行部１３１は、さらに、ストレージ装置１１０の記憶領域１１０ａに記憶されているデータの冗長構成が正しく保持されていることを確認する冗長パトロール処理を実行する。具体的には、パトロール処理実行部１３１は、管理情報記憶部１３３に記憶されている管理情報を参照して、ストレージ装置２１０に記憶されている冗長データから、ストレージ装置１１０に記憶されているデータと同一の内容の冗長データを特定する。次に、パトロール処理実行部１３１は、ストレージ装置１１０に記憶されているデータと、ストレージ装置２１０に記憶されている上記の特定した冗長データとを読み出す。

次に、パトロール処理実行部１３１は、読み出したデータと冗長データとを比較することによって、比較したデータと冗長データとが一致することを確認する。

【００６１】

ここで、パトロール処理実行部１３１は、パトロール処理および冗長パトロール処理を、ストレージ装置１１０の記憶領域１１０ａが測定単位領域からさらに複数に区分された領域である実行単位領域ごとにパトロールする。

【００６２】

また、パトロール処理実行部１３１は、パトロール処理として、ストレージ装置１１０との間で二重化されたストレージ装置２１０に記憶されているデータと同一の内容の冗長データをストレージ装置１１０内のデータの記憶領域１１０ａに対して、上書きで書き込む処理を実行する。このとき、パトロール処理実行部１３１は、管理情報記憶部１３３に記憶されている管理情報を参照して、ストレージ装置１１０に記憶されているデータと同一の内容の冗長データを特定する。次に、パトロール処理実行部１３１は、ストレージ装置１１０に記憶されているデータと同一の内容であって、ストレージ装置２１０に記憶されている、上記の特定した冗長データを読み出す。次に、パトロール処理実行部１３１は、読み出した冗長データをストレージ装置１１０内のデータの記憶領域１１０ａに対して、上書きで書き込む処理を実行する。

【００６３】

パトロール流量規制部１３２は、パトロール処理実行部１３１によって実行されているパトロール処理および冗長パトロール処理の速さであるパトロール流量を規制する。具体的には、パトロール流量規制部１３２は、パトロール性能測定部１３４によって測定されたパトロール性能に基づいてパトロール流量設定値を決定し、決定した当該パトロール流量設定値を超えないようにパトロール流量を規制する。このとき、パトロール流量規制部１３２は、ストレージ装置１１０の記憶領域をパトロール処理が実行される単位に区分した領域である、実行単位領域ごとにパトロール処理のパトロール流量を規制する。

【００６４】

また、パトロール流量規制部１３２は、時間帯に応じたパトロール流量の規制を行うこともできる。この場合、パトロール流量規制部１３２は、時刻情報取得部１３５によって取得された時刻情報が示す現在の時刻に基づいて、パトロール性能測定部１３４が保持するパトロール性能から現在の時刻が属する時間帯に対応するパトロール性能を選択する。次に、パトロール流量規制部１３２は、選択した当該パトロール性能に基づいてパトロール流量設定値を決定する。次に、パトロール流量規制部１３２は、決定したパトロール流量設定値を超えないようにパトロール流量を規制する。

【００６５】

管理情報記憶部１３３は、ストレージ装置１１０と、当該ストレージ装置１１０に記憶されているデータと同一内容の冗長データを記憶しているストレージ装置２１０とを対応付ける管理情報を記憶する。ストレージノード１００は、この管理情報に基づいて記憶しているデータの冗長データおよび冗長データが記憶されているストレージ装置を特定する。

【００６６】

パトロール性能測定部１３４は、ストレージ装置１１０における読み出し処理と書き込み処理との処理性能を示すパトロール性能を測定する。ここで、パトロール性能測定部１３４は、ストレージ装置１１０の記憶領域１１０ａが複数に区分された領域である測定単位領域ごとに、パトロール性能を測定する。

【００６７】

また、パトロール性能測定部１３４は、時刻情報取得部１３５によって取得された時刻情報が示す時刻が属する時間帯ごとに、測定したパトロール性能を複数種類保持する。

時刻情報取得部１３５は、現在の時刻を示す時刻情報を取得する。この時刻情報は、ストレージノード１００のＣＰＵ１０１が有するタイマによって生成されるが、これに限らず、ネットワーク１０に接続されたネットワーク時計、または他のノードから取得しても

よい。

【００６８】

このようなストレージノード１００によれば、パトロール処理実行部１３１により、ストレージ装置１１０の記憶領域が正常に稼動していることが確認される。このとき、パトロール流量規制部１３２により、パトロール処理実行部１３１によって実行されるパトロール処理の速さが規制される。この結果、パトロール処理で生じるストレージシステムへの負荷をコントロールすることによって、ストレージシステムに対するパトロール処理の影響を最小限に留めることができる。

【００６９】

次に、本実施の形態のパトロール性能の測定について説明する。

図６は、パトロール性能の測定の様子を示す図である。以下、図６に従って、本実施の形態のストレージノード１００によるパトロール性能の取得について説明する。

【００７０】

本実施の形態におけるパトロール処理の目的は、ストレージ装置１１０の記憶領域が正常に稼動していること、すなわち、ストレージ装置１１０の記憶領域に対して正常にアクセスできることを確認することである。そのため、本実施の形態のストレージノードは、パトロール処理において、ストレージ装置１１０内の記憶領域全体にわたってデータを読み出し、さらにそのデータを書き込む処理を実行する。

【００７１】

また、このパトロール処理がストレージシステムに与える負荷をコントロールするために、そのコントロールの基準とするべく、各ストレージ装置におけるパトロール処理の処理性能を示すパトロール性能を測定する。各ストレージノードでは、ストレージノードの起動時または定期的に、ストレージ装置全体またはストレージ装置の記憶領域のうちの特定の領域について、パトロール性能を測定する。また、各ストレージノードは、測定により得られたパトロール性能を、それぞれのストレージノードが有するＲＡＭ（例えば、ＲＡＭ１０２など）に格納する。

【００７２】

まず、ストレージノード１００によって管理されるＨＤＤ１１１およびストレージノード２００によって管理されるストレージ装置２１０を構成するＨＤＤ２１１の、それぞれの記憶領域全体のパトロール性能である、全体パトロール性能１１１ｐ，２１１ｐを測定する場合について説明する。この場合、ストレージノード１００は、例えば、ストレージシステムの起動時に、ＨＤＤ１１１の全体パトロール性能１１１ｐを測定する。ストレージノード２００は、同様に、ＨＤＤ２１１の全体パトロール性能２１１ｐを測定する。ストレージノード１００，２００は、全体パトロール性能１１１ｐ，２１１ｐのように、それぞれＨＤＤ１１１，２１１全体について一つのパトロール性能を保持する。そして、ストレージノード１００，２００は、それぞれのＨＤＤ１１１，２１１に対するパトロール処理の実行時において、この全体パトロール性能１１１ｐ，２１１ｐをパトロール流量の調整に用いる。

【００７３】

次に、ストレージ装置１１０を構成するＨＤＤ１１２の記憶領域を複数に分割した、それぞれの領域ごとのパトロール性能である領域別パトロール性能１１２ｐを測定する場合について説明する。この場合、ストレージノード１００は、例えば、ストレージシステムの起動時に、ＨＤＤ１１２の領域別パトロール性能１１２ｐを測定する。ストレージノード１００は、ＨＤＤ１１２について、領域別パトロール性能１１２ｐのように、複数に分割されたその領域ごとにパトロール性能を保持する。そして、ストレージノード１００は、ＨＤＤ１１２における各領域に対するパトロール処理の実行時において、この領域別パトロール性能１１２ｐを、パトロール流量の調整に用いる。

【００７４】

次に、ストレージ装置１１０を構成するＨＤＤ１１３の全体の時間帯ごとのパトロール性能である、時間帯別パトロール性能１１３ｐを測定する場合について説明する。この場

合、ストレージノード１００は、ＨＤＤ１１３全体についてのパトロール性能を、時間帯別パトロール性能１１３ｐのように、時間帯ごとに保持する。そして、ストレージノード１００は、ＨＤＤ１１３に対するパトロール処理の実行時の時間帯に合わせて、この時間帯別パトロール性能１１３ｐを、パトロール流量の調整に用いる。

【００７５】

次に、本実施の形態のストレージノード１００，２００，３００，４００で用いられるパトロール性能テーブルについて説明する。図７は、ストレージ装置ごとのパトロール性能の測定結果を格納したパトロール性能テーブルのデータ構造を示す図である。図８は、分割された領域ごとのストレージ装置のパトロール性能の測定結果を格納したパトロール性能テーブルのデータ構造を示す図である。図９は、時間帯ごとのストレージ装置のパトロール性能の測定結果を格納したパトロール性能テーブルのデータ構造を示す図である。

【００７６】

図７～図９に示すパトロール性能テーブルは、パトロール処理を行う際の処理を進める速さであるパトロール流量を調整するためのパトロール性能を示す情報である。パトロール情報は、ストレージ装置１１０，２１０，３１０，４１０が備える各ＨＤＤのデータ読み書きの性能特性を示す情報である。図６に示すように取得されたパトロール情報は、パトロール性能テーブル１９１，１９２，１９３としてテーブル化されて、ストレージノード１００のＲＡＭ１０２およびストレージノード２００，３００，４００のＲＡＭ（図示省略）に格納されている。

【００７７】

図７に示すパトロール性能テーブル１９１は、ストレージ装置ごとのパトロール性能である、全体パトロール性能（図６の全体パトロール性能１１１ｐを参照）の測定結果を格納するテーブルである。このパトロール性能テーブル１９１には、装置１９１ａ、シーケンシャル・アクセスの場合の読み出し処理１９１ｅおよび書き込み処理１９１ｆ、ならびにランダム・アクセスの場合の読み出し処理１９１ｇおよび書き込み処理１９１ｈの欄が設けられており、各欄の横方向に並べられた情報同士が互いに関連づけられている。

【００７８】

装置１９１ａの欄には、ストレージ装置１１０，２１０，３１０，および４１０に対して、それぞれが識別可能になるように割り当てられたストレージ装置の名称が格納される。

【００７９】

シーケンシャル・アクセスの場合の読み出し処理１９１ｅおよび書き込み処理１９１ｆの欄には、それぞれ、各ストレージ装置におけるシーケンシャル・アクセスによる読み出し処理および書き込み処理の際の全体パトロール性能を示す情報が格納される。シーケンシャル・アクセスによる読み出し処理および書き込み処理の際のパトロール性能は、データ転送速度、すなわち１秒間当りのデータ転送量（MB/s：MegaByte per second）によって示される。MB/sは、大きい値の方が高評価となる。

【００８０】

ランダム・アクセスの場合の読み出し処理１９１ｇおよび書き込み処理１９１ｈの欄には、それぞれ、各ストレージ装置におけるランダム・アクセスによる読み出し処理および書き込み処理の際の全体パトロール性能を示す情報が格納される。ランダム・アクセスによる読み出し処理および書き込み処理のパトロール性能は、処理能力、すなわち、１秒間当りの書き込み処理および読み出し処理の実行回数（IOPS：Input Output Per Second）によって示される。IOPSは、大きい値の方が高評価となる。

【００８１】

図８に示すパトロール性能テーブル１９２は、ストレージ装置における分割された領域ごとのパトロール性能である、領域別パトロール性能（図６の領域別パトロール性能１１２ｐを参照）の測定結果を格納するテーブルである。このパトロール性能テーブル１９２には、装置１９２ａ、領域１９２ｂ、シーケンシャル・アクセスの場合の読み出し処理１９２ｅおよび書き込み処理１９２ｆ、ならびにランダム・アクセスの場合の読み出し処理

１９２ｇおよび書き込み処理１９２ｈの欄が設けられており、各欄の横方向に並べられた情報同士が互いに関連づけられている。

【００８２】

装置１９２ａの欄には、パトロール性能テーブル１９１と同様、ストレージ装置１１０，２１０，３１０，および４１０に対して、それぞれが識別可能になるように割り当てられたストレージ装置の名称が格納される。

【００８３】

領域１９２ｂの欄には、各ストレージ装置における分割された領域の範囲を特定可能な情報が格納される。例えば、図８のパトロール性能テーブル１９２の一番上の行のように、領域１９２ｂが「０～１００［ｂｌｏｃｋ］」である場合には、その行には、ストレージ装置の領域０～１００［ｂｌｏｃｋ］の領域別パトロール性能が格納されることを示す。ここで、ｂｌｏｃｋは、各ＨＤＤにおけるデータの管理単位であり、ここでは、１ｂｌｏｃｋ＝１ＭＢを用いるが、必ずしもこれに限られない。

【００８４】

シーケンシャル・アクセスの場合の読み出し処理１９２ｅおよび書き込み処理１９２ｆの欄には、それぞれ、各ストレージ装置におけるシーケンシャル・アクセスによる読み出し処理および書き込み処理の際の領域別パトロール性能を示す情報が格納される。ランダム・アクセスの場合の読み出し処理１９２ｇおよび書き込み処理１９２ｈの欄には、それぞれ、各ストレージ装置におけるランダム・アクセスによる読み出し処理および書き込み処理の際の領域別パトロール性能を示す情報が格納される。

【００８５】

図９に示すパトロール性能テーブル１９３は、各ストレージ装置における時間帯ごとのパトロール性能である、時間帯別パトロール性能（図６の時間帯別パトロール性能１１３ｐを参照）の測定結果を格納するテーブルである。このパトロール性能テーブル１９３には、装置１９３ａ、対象時間１９３ｃ、測定日１９３ｄ、シーケンシャル・アクセスの場合の読み出し処理１９３ｅおよび書き込み処理１９３ｆ、ならびにランダム・アクセスの場合の読み出し処理１９３ｇおよび書き込み処理１９３ｈの欄が設けられており、各欄の横方向に並べられた情報同士が互いに関連づけられている。

【００８６】

装置１９３ａの欄には、パトロール性能テーブル１９１と同様、ストレージ装置１１０，２１０，３１０，および４１０に対して、それぞれが識別可能になるように割り当てられたストレージ装置の名称が格納される。

【００８７】

対象時間１９３ｃの欄には、各ストレージ装置の時間帯別パトロール性能が測定された時間帯を特定可能な情報が格納される。例えば、図９のパトロール性能テーブル１９３の一番上の行のように、対象時間１９３ｃが「０～３時」である場合には、その行には、０時から３時に測定されたパトロール性能が格納されることを示す。

【００８８】

測定日１９３ｄの欄には、各ストレージ装置の時間帯別パトロール性能が測定された日付を特定可能な情報が格納される。例えば、図９のパトロール性能テーブル１９３の一番上の行のように、測定日１９３ｄが「１０／１」である場合には、その行の時間帯別パトロール性能は、１０月１日に測定されたものであることを示す。

【００８９】

シーケンシャル・アクセスの場合の読み出し処理１９３ｅおよび書き込み処理１９３ｆの欄には、それぞれ、各ストレージ装置におけるシーケンシャル・アクセスによる読み出し処理および書き込み処理の際の時間帯別パトロール性能を示す情報が格納される。ランダム・アクセスの場合の読み出し処理１９３ｇおよび書き込み処理１９３ｈの欄には、それぞれ、各ストレージ装置におけるランダム・アクセスによる読み出し処理および書き込み処理の際の時間帯別パトロール性能を示す情報が格納される。

【００９０】

ストレージノード１００，２００，３００，４００は、図１１に示すパトロール処理の実行時、および図１３に示す冗長パトロール処理の実行時に、対応付けられているストレージ装置１１０，２１０，３１０，４１０からパトロール性能を収集し、それぞれが備えるＲＡＭのパトロール性能テーブル１９１，１９２，１９３に格納する。

【００９１】

また、ストレージ装置１１０，２１０，３１０，４１０のパトロール性能として、例えば、応答時間など、その他の性能を用いてもよい。

ここで、パトロール性能とパトロール流量との関係について説明する。図７～図９に示すパトロール性能テーブル１９１，１９２，１９３に格納されているパトロール性能に基づいて、例えば、以下の（例１）～（例３）に示すようなポリシーに従い、ストレージノード１００により、パトロール流量が決定される。ここでは、“Seq-R”は、シーケンシャル・アクセスによる読み出しであり、“Seq-W”は、シーケンシャル・アクセスによる書き込みである。また、“Ran-R”は、ランダム・アクセスによる読み出しであり、“Ran-W”は、ランダム・アクセスによる書き込みである。

【００９２】

（例１）「アクセスに影響を与えない範囲で可能なかぎりパトロールを行う」場合

パトロール処理の実行開始前の１分間の平均アクセスが、

Seq-R 10[MB/s]
Seq-W 0[MB/s]
Ran-R 100[IOPS]
Ran-W 0[IOPS]

であり、ストレージ装置１１０の最大性能が、

Seq-R 50[MB/s]
Seq-W 50[MB/s]
Ran-R 500[IOPS]
Ran-W 500[IOPS]

である場合、シーケンシャル・アクセスの読み出しで、ストレージ装置１１０の最大性能の２０％を占めており、ランダム・アクセスの読み出しで、最大性能の２０％を占めている。このため、パトロール処理に対しては、ストレージ装置１１０の最大性能の６０％を割り当てることができる。

【００９３】

ここで、パトロール処理では、シーケンシャル・アクセスの読み出しおよび書き込みが平行して行われる。このため、パトロール流量は、それぞれ、シーケンシャル・アクセスの読み出しおよび書き込みに対して割り当て可能な６０％の二分の一となり、１５[MB/s]となる。

【００９４】

（例２）「ランダム・アクセスが性能の５０％を超えているときには、アクセスに影響を与えない範囲で可能なかぎりパトロールを行い、５０％以下の場合にはパトロール処理に性能の２０％を費やすようにする」場合

パトロール処理の実行開始前の１分間の平均アクセスが、

Seq-R 20[MB/s]
Seq-W 0[MB/s]
Ran-R 100[IOPS]
Ran-W 100[IOPS]

であり、性能値が、

Seq-R 50[MB/s]
Seq-W 50[MB/s]
Ran-R 500[IOPS]
Ran-W 500[IOPS]

である場合、ランダム・アクセスは、読み出しおよび書き込みにおいて、それぞれパトロ

ール性能の２０％を占めている。このランダム・アクセスの読み出しおよび書き込みは、合わせてストレージ装置１１０の最大性能の４０％を占めている。すなわち、この時点のランダム・アクセスは、ストレージ装置１１０の最大性能の５０％以下である。

【００９５】

ここで、パトロール処理では、シーケンシャル・アクセスの読み出しおよび書き込みが平行して行われる。このため、ディスク性能の２０％を割り当てると、１０[MB/s]となることから、パトロール流量は、その二分の一の５[MB/s] となる。

【００９６】

（例３）「アクセス状況に関係なく、常に１[MB/s]でパトロール処理を行う」場合

パトロール流量は、常に１[MB/s]となる。

次に、ストレージ装置の記憶領域において、パトロール性能が測定される分割単位およびパトロール処理が実行される分割単位の関係について説明する。図１０は、パトロール処理の実行時におけるストレージ装置の領域の分割を示す図である。ここでは、ストレージ装置１１０の有する記憶領域を一体として、記憶領域１１６として説明する。

【００９７】

図１０に示す、ストレージ装置１１０の記憶領域１１６は、パトロール性能の測定の際の分割単位である複数の測定単位１１６ａ，１１６ｂ，１１６ｃ，１１６ｄ，・・・，１１６ｎに分割されている。図８に示す領域別パトロール性能は、この測定単位１１６ａ，１１６ｂ，１１６ｃ，１１６ｄ，・・・，１１６ｎごとに測定される。

【００９８】

また、記憶領域１１６は、さらに、パトロール処理が実行される際の分割単位である複数の実行単位１１６ａ１，１１６ａ２，・・・，１１６ａｎに分割されている。後述する図１１に示すパトロール処理および図１４に示す冗長パトロール処理は、この実行単位１１６ａ１，１１６ａ２，・・・，１１６ａｎごとに処理が進められ、パトロール流量の規制が行われる。

【００９９】

次に、本実施の形態のストレージノードにおいて実行される処理の手順について説明する。まず、本実施の形態のストレージノードにおいて実行されるパトロール処理について説明する。図１１は、パトロール処理の手順を示すフローチャートである。

【０１００】

本実施の形態のストレージノード１００では、例えば、ストレージシステムの起動時、ストレージノードまたはストレージ装置の起動時、所定の時間の経過時などに、このパトロール処理の実行が開始される。なお、パトロール処理は、ストレージ装置について記憶領域に異常がないかチェックするためにパトロールを行う処理である。

【０１０１】

以下に、図１１に従って、ストレージノード１００が、ストレージ装置１１０についてパトロールを行う場合について説明する。ここでは、ストレージノード１００の起動時において、ストレージ装置１１０の全領域についてパトロールが行われる場合に基づいて説明する。ストレージノード１００の起動時において、ストレージノード１００は、ストレージ装置１１０の全領域が正常に読み書き可能であるか否かを診断するパトロール処理を実行するために、パトロール処理を呼び出す。

【０１０２】

［ステップＳ１１］ストレージノード１００は、パトロール流量を設定する。このパトロール流量（図６参照）は、上記のように、パトロール処理が行われる速さを調整するための条件である。ここで、ストレージノード１００は、パトロール性能として、全体パトロール性能（例えば、図６に示す全体パトロール性能１１１ｐ，２１１ｐ）、領域別パトロール性能（例えば、図６に示す領域別パトロール性能１１２ｐ）、および時間帯別パトロール性能（例えば、図６に示す時間帯別パトロール性能１１３ｐ）を用いることができる。ストレージノード１００が各ストレージ装置に対していずれのパトロール性能を用い

るかについては、各ストレージ装置のハードウェア特性、使用状況に合わせて予め管理者などにより設定される。

【０１０３】

例えば、記憶領域によってパトロール性能が大きく異なるストレージ装置については、各領域ごとのパトロール性能を示す領域別パトロール性能を用いる。また、時間帯によって使用状況に偏りがあるストレージ装置については、時間帯ごとのパトロール性能を示す時間帯別パトロール性能を用いる。また、パトロール性能が変化しないストレージ装置については、全体パトロール性能を用いる。ストレージノード１００は、これらのパトロール性能を予め測定しておき、パトロール処理の実行時に処理の実行対象のストレージ装置が有するＨＤＤのパトロール性能を取得する。そして、ストレージノード１００は、取得したパトロール性能を、そのストレージ装置における各ＨＤＤのパトロール流量として設定する。

【０１０４】

［ステップＳ１２］ストレージノード１００は、ストレージ装置１１０について記憶領域の実行単位ごとにパトロールするパトロール単位実行処理を実行する。詳しくは、図１２に示すパトロール単位実行処理において後述する。また、ストレージノード１００は、後述する指定時間の起算点とするために、パトロール単位実行処理の実行開始時に、時刻情報を取得する。

【０１０５】

［ステップＳ１３］ストレージノード１００は、ステップＳ１１において設定されたパトロール流量以下でパトロールが実行されているか否かを判定する。ストレージノード１００は、設定されたパトロール流量を超えていれば、ステップＳ１４に処理を進める一方、設定されたパトロール流量以下であれば、ステップＳ１５に処理を進める。

【０１０６】

ステップＳ１３において、具体的には、ストレージノード１００は、実行単位（図１０参照）ごとのパトロール処理の時間間隔が、設定された指定時間以下であるか否かを判定する。

【０１０７】

ここで、指定時間は、パトロール処理がストレージシステムに与える負荷を調整するべく、実行単位ごとのパトロール処理の時間間隔が一定時間以上になるように規制するための値である。この指定時間は、ＨＤＤのパトロール性能が高ければパトロール流量を増加（パトロール処理の速さの制限を緩和）させ、パトロール性能が低ければパトロール流量を低下（パトロール処理の速さの制限を強化）させるものを用いることができる。

【０１０８】

本実施の形態では、指定時間を、以下の関係式を用いて算出する。

指定時間[s]＝実行単位[MB]／パトロール性能[MB/s]　　（１）

具体的には、例えば、あるＨＤＤの全体パトロール性能が１０[MB/s]の場合において、このＨＤＤの実行単位が１[MB]のときは、指定時間は、１[MB]／１０[MB/s]＝０．１[s]となる。

【０１０９】

本実施の形態のストレージノード１００は、以上のように指定時間をこのように算出する。そして、ストレージノード１００は、直近のパトロール単位実行処理（Ｓ１２）の開始時点から起算して指定時間が経過していれば、「設定流量以下である」と判定する一方、前回のパトロール処理の開始時点から指定時間が経過していなければ、「設定流量を超えた」と判定する。

【０１１０】

［ステップＳ１４］ストレージノード１００は、直近のパトロール単位実行処理の開始時点から起算して指定時間が経過するまでステップＳ１４において処理を待機する。ストレージノード１００は、指定時間が経過した後、ステップＳ１５に処理を進める。

【０１１１】

具体的には、ストレージノード１００は、前回のパトロール処理が終了してから一定時間（例えば、０．１秒）経過するまで、ステップＳ１４において処理を待機する。

［ステップＳ１５］ストレージノード１００は、ステップＳ１２のパトロール単位実行処理が、ストレージ装置１１０のすべての実行単位について実行されたか否かを判定する。ストレージノード１００は、すべての実行単位について実行されていなければ、ステップＳ１２に処理を進める一方、すべての実行単位について実行されていれば、処理を終了する。

【０１１２】

次に、上記のパトロール処理（図１１参照）のステップＳ１２において実行されるパトロール単位実行処理の詳細について説明する。図１２は、パトロール単位実行処理の手順を示すフローチャートである。ここでは、図１１に示すパトロール処理と同様、ストレージノード１００の起動時において、ストレージ装置１１０の全領域についてパトロールが行われる場合に基づいて説明する。

【０１１３】

［ステップＳ２１］ストレージノード１００は、ストレージ装置１１０の記憶領域における現在のパトロール対象の実行単位に格納されているデータを読み出す。

［ステップＳ２２］ストレージノード１００は、ステップＳ２１における読み出しに成功したか否かを判定する。ストレージノード１００は、読み出しに成功していれば、ステップＳ２３に処理を進める一方、読み出しに失敗していれば、ステップＳ２４に処理を進める。

【０１１４】

［ステップＳ２３］ストレージノード１００は、ステップＳ２１において読み出したデータを、ステップＳ２１においてデータを読み出した記憶領域に書き込む。その後、ストレージノード１００は、パトロール単位実行処理を終了してパトロール処理に復帰する。

【０１１５】

［ステップＳ２４］ストレージノード１００は、ステップＳ２１において読み出したデータと対応する冗長データを、冗長データが記憶されているストレージ装置から読み出して、ステップＳ２１においてデータを読み出した記憶領域に書き込む。その後、ストレージノード１００は、パトロール単位実行処理を終了してパトロール処理に復帰する。

【０１１６】

以上のようにして、本実施の形態のストレージノード１００は、パトロール処理を実行することにより、ストレージ装置１１０の記憶領域が正常に稼動していること、すなわち、ストレージ装置１１０の記憶領域に正常に読み書きできることを確認する。また、パトロール処理を自律的に調整して実行することで、パトロール処理に基づくストレージシステムの負荷を規制する。この結果、パトロール処理の実行に基づく負荷による、ストレージ装置へのアクセスに対する影響を抑えつつ、ストレージシステムの信頼性を高めることができる。

【０１１７】

次に、本実施の形態のストレージノード１００で実行される冗長パトロールについて説明する。本実施の形態のストレージノード１００では、上記のパトロール処理に加えて、冗長パトロール処理が実行される。この冗長パトロール処理は、ストレージ装置１１０に記憶されているデータをバックアップするためにそのデータと同一の内容である冗長データがストレージ装置１１０と二重化されている他のストレージ装置（例えば、ストレージ装置２１０）に記憶されている場合には、ストレージ装置１１０に記憶されているデータと、そのデータに対応する冗長データの整合性について確認する処理である。

【０１１８】

本実施の形態のストレージノード１００では、冗長パトロール処理をストレージノード１００の起動時および定期的に実行する。このとき、冗長パトロール処理は、ストレージシステムに対する負荷を調整するために、冗長パトロール性能に基づいて実行される。

【０１１９】

次に、本実施の形態の冗長パトロール性能の測定について説明する。

図１３は、冗長パトロール性能の測定の様子を示す図である。以下、図１３に従って、本実施の形態のストレージノード１００による冗長パトロール性能の取得について説明する。

【０１２０】

本実施の形態における冗長パトロール処理の目的は、ストレージ装置１１０の記憶領域に記憶されているデータが、対応する冗長データと整合し、冗長構成が維持されていること、すなわち、ストレージ装置１１０の記憶されているデータが二重化されているストレージ装置によって正常にバックアップされていることを確認することである。そのため、本実施の形態のストレージノードは、冗長パトロール処理によって、ストレージ装置１１０内の記憶領域全体にわたってデータを読み出し、さらにそのデータを書き込む処理を実行する。

【０１２１】

また、パトロール処理と同様、この冗長パトロール処理がストレージシステムに与える負荷をコントロールするために、そのコントロールの基準とするべく、冗長パトロール処理の処理性能を示すパトロール性能である冗長パトロール性能を測定する。このとき、冗長構成をとるストレージ装置を管理するストレージノード間で、協調して冗長パトロール処理を進める必要があるため、パトロール性能としてノード間の冗長構成も考慮に入れた値である冗長パトロール性能を測定し、それを冗長パトロール処理における処理の速さ（冗長パトロール流量）の調整に用いる。

【０１２２】

各ストレージノードでは、ストレージノードの起動時または定期的に、ストレージ装置全体またはストレージ装置の記憶領域のうちの特定の領域について、二重化された他方のストレージ装置を管理するストレージノードと協調して、冗長パトロール性能を測定する。また、各ストレージノードは、測定により得られた冗長パトロール性能を、それぞれのストレージノードが有するＲＡＭ（例えば、ＲＡＭ１０２など）に格納する。

【０１２３】

図１３は、ストレージノード１００が管理するストレージ装置１１０のＨＤＤ１１１とストレージノード２００が管理するストレージ装置２１０のＨＤＤ２１１とが二重化（ＲＡＩＤ１）構成となっているときの冗長パトロール性能５１１ｐを測定する場合の測定の様子について図示したものである。この場合、ストレージノード１００，２００は、冗長パトロール性能５１１ｐのように、ＨＤＤ１１１，２１１の二つのＨＤＤの組ごとに一つの冗長パトロール性能値を保持する。

【０１２４】

次に、本実施の形態のストレージノードにおいて実行される冗長パトロール処理について説明する。冗長パトロール処理は、ストレージ装置に記憶されているデータと、冗長構成のストレージ装置に記憶されている冗長データとの整合性をチェックするためにパトロールを行う処理である。図１４は、冗長パトロール処理の手順を示すフローチャートである。

【０１２５】

以下に、図１４に従って、ストレージノード１００が、ストレージノード２００とともに、ストレージ装置１１０および冗長構成のストレージ装置２１０について冗長パトロールを行う場合について説明する。この冗長パトロール処理は、二重化されたストレージ装置のうちの一方のストレージ装置（例えば、ストレージ装置１１０）を管理するストレージノード（例えば、ストレージノード１００）が主導して実行するが、他方のストレージ装置（例えば、ストレージ装置２１０）を管理するストレージノード（例えば、ストレージノード２００）と協調して処理が進められる。ここでは、ストレージノード１００の起動時におけるパトロール処理が実行された後、ストレージ装置１１０の全領域、およびこれに冗長構成として関係付けられているストレージ装置２１０の全領域についてパトロールが行われる場合に基づいて説明する。

【０１２６】
ストレージノード１００の起動時においてパトロール処理が実行された後、ストレージノード１００は、ストレージ装置１１０の全領域が、ストレージ装置１１０のデータと冗長構成として関係付けられたストレージ装置２１０のデータとの間で、整合性を有しているか否かを診断する冗長パトロール処理を実行するために、冗長パトロール処理を呼び出す。
【０１２７】
［ステップＳ３１］ストレージノード１００は、パトロール処理を実行する二重化されたＨＤＤの組に対応する冗長パトロール流量（図１３参照）を設定する。
［ステップＳ３２］ストレージノード１００は、ストレージ装置１１０について記憶領域の実行単位ごとにパトロールするとともに、ストレージ装置２１０を管理するストレージノードがストレージ装置２１０をパトロールしたパトロール結果を受信して、両者の整合性をチェックする冗長パトロール単位実行処理を実行する。詳しくは、図１５に示す冗長パトロール単位実行処理において後述する。また、ストレージノード１００は、後述する指定時間の起算点とするために、冗長パトロール単位実行処理の実行開始時に、時刻情報を取得する。
【０１２８】
［ステップＳ３３］ストレージノード１００は、パトロール処理のステップＳ１３と同様に、ステップＳ３１において設定されたパトロール流量以下で冗長パトロールが実行されているか否かを判定する。ストレージノード１００は、設定されたパトロール流量を超えていれば、ステップＳ３４に処理を進める一方、設定されたパトロール流量以下であれば、ステップＳ３５に処理を進める。
【０１２９】
［ステップＳ３４］ストレージノード１００は、直近の冗長パトロール単位実行処理（ステップＳ３２）の開始時点から起算して指定時間が経過するまでステップＳ３４において処理を待機する。ストレージノード１００は、指定時間が経過した後、ステップＳ３５に処理を進める。
【０１３０】
［ステップＳ３５］ストレージノード１００は、ステップＳ３２の冗長パトロール単位実行処理が、ストレージ装置１１０のすべての実行単位について実行されたか否かを判定する。ストレージノード１００は、すべての実行単位について実行されていなければ、ステップＳ３２に処理を進める一方、すべての実行単位について実行されていれば、処理を終了する。
【０１３１】
次に、上記の冗長パトロール処理（図１４参照）のステップＳ３２において実行される冗長パトロール単位実行処理の詳細について説明する。図１５は、冗長パトロール単位実行処理の手順を示すシーケンス図である。ここでは、ストレージノード１００の起動時におけるパトロール処理が実行された後、ストレージ装置１１０の全領域、およびこれに冗長構成として関係付けられているストレージ装置２１０の全領域についてパトロールが行われる場合に基づいて説明する。
【０１３２】
［ステップＳ４１］ストレージノード１００は、冗長構成としてストレージ装置１１０と関係付けられたストレージ装置２１０を管理するストレージノード２００に対して、ストレージ装置２１０に対する冗長パトロール処理の開始を要求する冗長パトロール要求を送信する。このとき、ストレージノード１００は、併せて、ストレージノード２００に対して、冗長パトロール処理の対象であるストレージ装置２１０の実行単位を特定するための情報を送信する。
【０１３３】
［ステップＳ４２］ストレージノード１００は、ストレージ装置１１０の記憶領域における現在のパトロール対象の実行単位に格納されているデータを読み出す。

［ステップＳ４３］ストレージノード１００は、ストレージノード２００から送信された、ストレージ装置２１０に記憶されていた冗長データを受信する。

【０１３４】

［ステップＳ４４］ストレージノード１００は、ステップＳ４２において読み出したデータとステップＳ４３で受信したデータの整合性をチェックする。

［ステップＳ４５］ストレージノード１００は、ステップＳ４４において行った整合性のチェック結果を、ストレージノード２００に送信する。

【０１３５】

［ステップＳ４６］ストレージノード１００は、ステップＳ４４において行った整合性のチェック結果を、ストレージノード１００のＲＡＭ１０２に保存する。その後、ストレージノード１００は、冗長パトロール単位実行処理を終了して冗長パトロール処理に復帰する。

【０１３６】

［ステップＳ５１］ストレージノード２００は、ステップＳ４１においてストレージノード１００から送信された冗長パトロール要求を受信する。このとき、ストレージノード２００は、併せて送信された情報から、冗長パトロール処理の対象であるストレージ装置２１０の実行単位を特定する。

【０１３７】

［ステップＳ５２］ストレージノード２００は、ストレージ装置２１０の記憶領域におけるステップＳ５１において特定された実行単位に格納されているデータを読み出す。このデータは、ストレージ装置１１０に記憶されている、冗長パトロールの対象であるデータの冗長データである。

【０１３８】

［ステップＳ５３］ストレージノード２００は、ステップＳ５２においてストレージ装置２１０から読み出した冗長データを、ストレージノード１００に対して送信する。

［ステップＳ５４］ストレージノード２００は、ステップＳ４５においてストレージノード１００から送信された整合性のチェック結果を受信する。

【０１３９】

［ステップＳ５５］ストレージノード２００は、ステップＳ５４において受信した整合性のチェック結果を、ストレージノード２００のＲＡＭ（図示省略）に保存する。その後、ストレージノード２００は、処理を終了する。

【０１４０】

以上のようにして、本実施の形態のストレージノードは、冗長パトロール処理を実行することにより、ストレージノード間で冗長構成をとっている場合に、各ノードのデータが正しく冗長構成を保持しているか否かを診断することができる。

【０１４１】

以上、本発明のストレージ管理装置、ストレージ管理プログラムおよびストレージシステムを、図示の実施の形態に基づいて説明したが、上記については単に本発明の原理を示すものである。本発明は上記に示し、説明した正確な構成および応用例に限定されるものではなく、さらに、多数の変形、変更が当業者にとって可能であり、対応するすべての変形例および均等物は、添付の請求項およびその均等物による本発明の範囲とみなされ、各部の構成は同様の機能を有する任意の構成のものに置換することができる。また、本発明に他の任意の構成物や行程が付加されてもよい。また、本発明は前述した実施の形態のうちの任意の２以上の構成（特徴）を組み合わせたものであってもよい。

【０１４２】

なお、上記の処理機能は、コンピュータによって実現することができる。その場合、ストレージノード１００，２００，３００，４００、管理ノード３０、端末装置２１，２２，２３、コントロールノード５００およびアクセスノード６００が有すべき機能の処理内容を記述したプログラムが提供される。そのプログラムをコンピュータで実行することにより、上記処理機能がコンピュータ上で実現される。

【０１４３】
処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体には、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリなどがある。磁気記録装置には、ＨＤＤ、フレキシブルディスク（ＦＤ）、磁気テープ（ＭＴ）などがある。光ディスクには、ＤＶＤ（Digital Versatile Disc）、ＤＶＤ－ＲＡＭ、ＣＤ－ＲＯＭ（Compact Disc - Read Only Memory）、ＣＤ－Ｒ（Recordable）／ＲＷ（ReWritable）などがある。光磁気記録媒体には、ＭＯ（Magneto - Optical disk）などがある。
【０１４４】
上記プログラムを流通させる場合には、例えば、そのプログラムが記録されたＤＶＤ、ＣＤ－ＲＯＭなどの可搬型記録媒体が販売される。また、プログラムをサーバコンピュータに格納しておき、ネットワークを通じて、サーバコンピュータから他のコンピュータにそのプログラムを転送することもできる。
【０１４５】
上記プログラムを実行するコンピュータは、例えば、可搬型記録媒体に記録されたプログラム若しくはサーバコンピュータから転送されたプログラムを、自己の記憶装置に格納する。そして、コンピュータは、自己の記憶装置からプログラムを読み取り、プログラムに従った処理を実行する。なお、コンピュータは、可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することもできる。また、コンピュータは、サーバコンピュータからプログラムが転送されるごとに、逐次、受け取ったプログラムに従った処理を実行することもできる。
【０１４６】
以上の実施の形態に関し、さらに以下の付記を開示する。
（付記１）　データをネットワークで接続された複数のストレージ装置に分散して記憶するストレージシステムにおけるストレージ装置を管理するストレージ管理装置において、
前記ストレージ装置の記憶領域が正常に稼動していることを確認するパトロール処理を実行するパトロール処理実行手段と、
前記パトロール処理実行手段によって実行されている前記パトロール処理の速さであるパトロール流量を規制するパトロール流量規制手段と、
を有することを特徴とするストレージ管理装置。
【０１４７】
（付記２）　前記ストレージ装置と、当該ストレージ装置に記憶されている前記データと同一内容の冗長データを記憶している二重化ストレージ装置とを対応付ける管理情報を記憶する管理情報記憶手段を有し、
前記パトロール処理実行手段は、さらに、前記パトロール処理において、前記管理情報記憶手段に記憶されている前記管理情報を参照して、前記ストレージ装置に記憶されている前記データと、前記二重化ストレージ装置に記憶されている前記データと同一内容の前記冗長データとを読み出し、読み出した前記データと前記冗長データとを比較することによって、比較した前記データと前記冗長データとが一致することを確認することを特徴とする付記１記載のストレージ管理装置。
【０１４８】
（付記３）　前記ストレージ装置における読み出し処理と書き込み処理との処理性能を示すパトロール性能を測定するパトロール性能測定手段を有し、
前記パトロール流量規制手段は、前記パトロール性能測定手段によって測定された前記パトロール性能に基づいて前記パトロール流量設定値を決定し、決定した当該パトロール流量設定値を超えないように前記パトロール流量を規制することを特徴とする付記１記載のストレージ管理装置。
【０１４９】
（付記４）　前記パトロール性能測定手段は、前記ストレージ装置の記憶領域が複数に

区分された領域である測定単位領域ごとに、前記パトロール性能を測定し、
前記パトロール処理実行手段は、前記パトロール処理を、前記ストレージ装置の記憶領域が前記測定単位領域からさらに複数に区分された領域である実行単位領域ごとに実行し、
前記パトロール流量規制手段は、前記実行単位領域ごとに前記パトロール処理の前記パトロール流量を規制することを特徴とする付記１記載のストレージ管理装置。
【０１５０】
（付記５）　現在の時刻を示す時刻情報を取得する時刻情報取得手段を有し、
前記パトロール性能測定手段は、前記時刻情報取得手段によって取得された前記時刻情報が示す時刻が属する時間帯ごとに、測定した前記パトロール性能を複数種類保持し、
前記パトロール流量規制手段は、前記時刻情報取得手段によって取得された前記時刻情報が示す前記現在の時刻に基づいて、前記パトロール性能測定手段が保持する前記パトロール性能から前記現在の時刻が属する時間帯に対応する前記パトロール性能を選択し、選択した当該パトロール性能に基づいて前記パトロール流量設定値を決定し、決定した当該パトロール流量設定値を超えないように前記パトロール流量を規制することを特徴とする付記３記載のストレージ管理装置。
【０１５１】
（付記６）　前記ストレージ装置と、当該ストレージ装置に記憶されている前記データと同一内容の冗長データを記憶している二重化ストレージ装置とを対応付ける管理情報を記憶する管理情報記憶手段を有し、
前記パトロール処理実行手段は、前記パトロール処理として、前記管理情報記憶手段に記憶されている前記管理情報を参照して、前記二重化ストレージ装置に記憶されている前記データと同一内容の前記冗長データとを読み出し、読み出した前記冗長データを前記ストレージ装置内の前記データの記憶領域に対して上書きで書き込む処理を実行することを特徴とする付記１記載のストレージ管理装置。
【０１５２】
（付記７）　コンピュータに、データをネットワークで接続された複数のストレージ装置に分散して記憶するストレージシステムにおけるストレージ装置の管理処理を実行させるストレージ管理プログラムにおいて、
前記コンピュータを、
前記ストレージ装置の記憶領域が正常に稼動していることを確認するパトロール処理を実行するパトロール処理実行手段、
前記パトロール処理実行手段によって実行されている前記パトロール処理の速さであるパトロール流量を規制するパトロール流量規制手段、
として機能させることを特徴とするストレージ管理プログラム。
【０１５３】
（付記８）　データをネットワークで接続された複数のストレージ装置に分散して記憶し、前記ストレージ装置を管理するストレージ管理装置を有するストレージシステムにおいて、
前記ストレージ装置と、当該ストレージ装置に記憶されている前記データと同一内容の冗長データを記憶している二重化ストレージ装置とを対応付ける管理情報を記憶する管理情報記憶手段を有し、
前記ストレージ管理装置は、
前記ストレージ装置の記憶領域が正常に稼動していることを確認するパトロール処理を実行するパトロール処理実行手段と、
前記パトロール処理実行手段によって実行されている前記パトロール処理の速さであるパトロール流量を規制するパトロール流量規制手段と、
を有し、
前記パトロール処理実行手段は、さらに、前記パトロール処理において、前記管理情報記憶手段に記憶されている前記管理情報を参照して、前記ストレージ装置に記憶されてい

る前記データと、前記二重化ストレージ装置に記憶されている前記データと同一内容の前記冗長データとを読み出し、読み出した前記データと前記冗長データとを比較することによって、比較した前記データと前記冗長データとが一致することを確認することを特徴とするストレージシステム。

【図面の簡単な説明】

【０１５４】

【図１】本実施の形態の概要を示す図である。

【図２】本実施の形態のシステム構成を示す図である。

【図３】ストレージノードのハードウェア構成を示す図である。

【図４】論理ボリュームのデータ構造を示す図である。

【図５】ストレージノードの機能を示すブロック図である。

【図６】パトロール性能の測定の様子を示す図である。

【図７】ストレージ装置ごとのパトロール性能の測定結果を格納したパトロール性能テーブルのデータ構造を示す図である。

【図８】分割された領域ごとのストレージ装置のパトロール性能の測定結果を格納したパトロール性能テーブルのデータ構造を示す図である。

【図９】時間帯ごとのストレージ装置のパトロール性能の測定結果を格納したパトロール性能テーブルのデータ構造を示す図である。

【図１０】パトロール処理の実行時におけるストレージ装置の領域の分割を示す図である。

【図１１】パトロール処理の手順を示すフローチャートである。

【図１２】パトロール単位実行処理の手順を示すフローチャートである。

【図１３】冗長パトロール性能の測定の様子を示す図である。

【図１４】冗長パトロール処理の手順を示すフローチャートである。

【図１５】冗長パトロール単位実行処理の手順を示すシーケンス図である。

【符号の説明】

【０１５５】

１　ストレージ管理装置

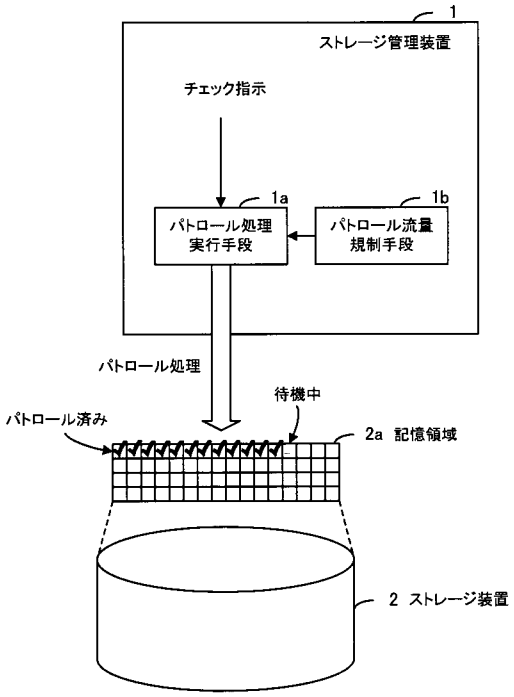
１ａ　パトロール処理実行手段

１ｂ　パトロール流量規制手段

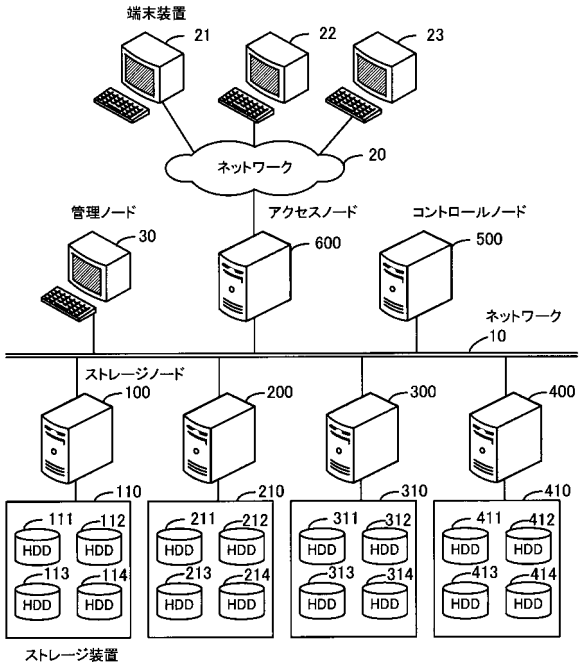
２　ストレージ装置

２ａ　記憶領域

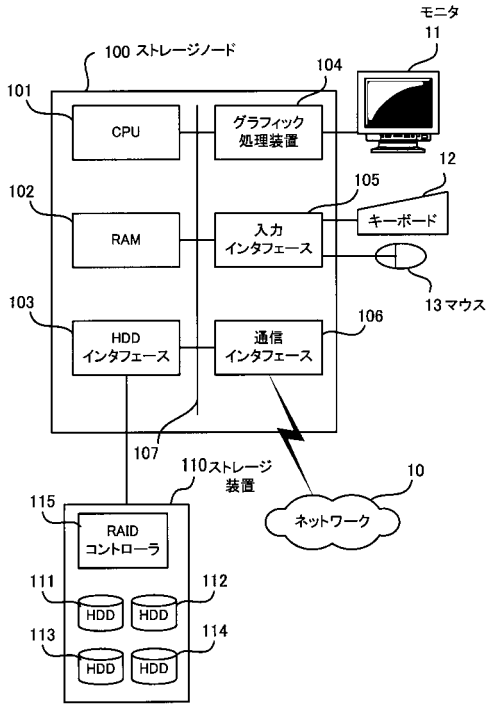
【図１】


1
ストレージ管理装置
チェック指示
1a
パトロール処理
実行手段
1b
パトロール流量
規制手段
パトロール処理
待機中
パトロール済み
2a 記憶領域
2 ストレージ装置

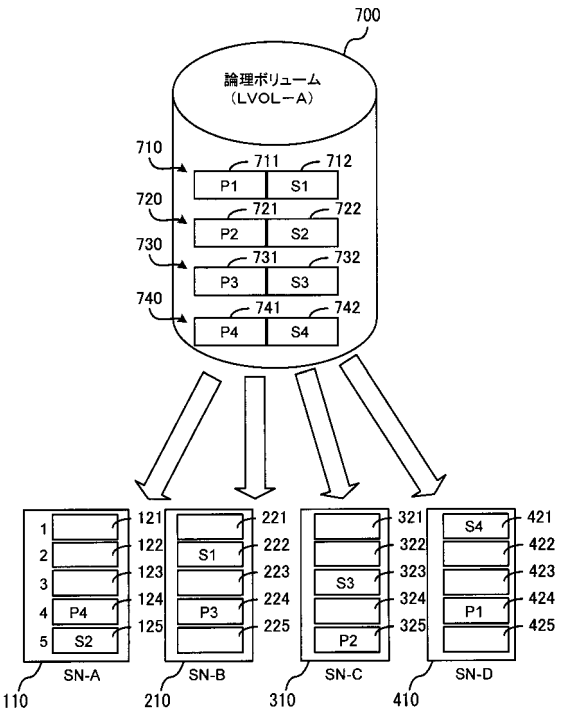
【図２】


端末装置
21
22
23
ネットワーク
20
管理ノード
30
アクセスノード
600
コントロールノード
500
ネットワーク
10
ストレージノード
100
200
300
400
110
210
310
410
111
112
113
114
211
212
213
214
311
312
313
314
411
412
413
414
HDD
ストレージ装置

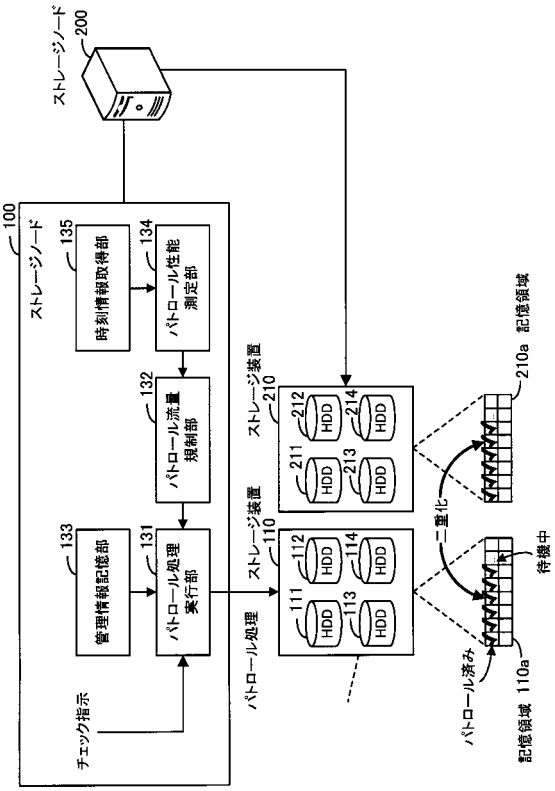
【図３】


モニタ
11
100 ストレージノード
104
101
CPU
グラフィック
処理装置
12
105
102
RAM
入力
インタフェース
キーボード
13 マウス
103
HDD
インタフェース
通信
インタフェース
106
107
110ストレージ
装置
10
ネットワーク
115
RAID
コントローラ
111
112
113
114
HDD

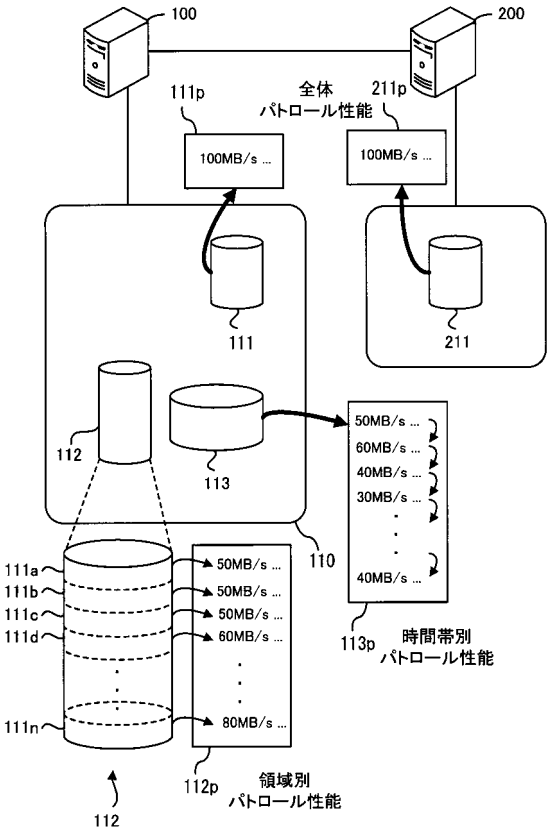
【図４】


700
論理ボリューム
(LVOL-A)
710
711
712
P1
S1
720
721
722
P2
S2
730
731
732
P3
S3
740
741
742
P4
S4
121
122
123
124
125
221
222
223
224
225
321
322
323
324
325
421
422
423
424
425
1
2
3
4
5
S1
P3
S3
P2
S4
P1
SN-A
SN-B
SN-C
SN-D
110
210
310
410

【図５】



【図６】



【図７】

191

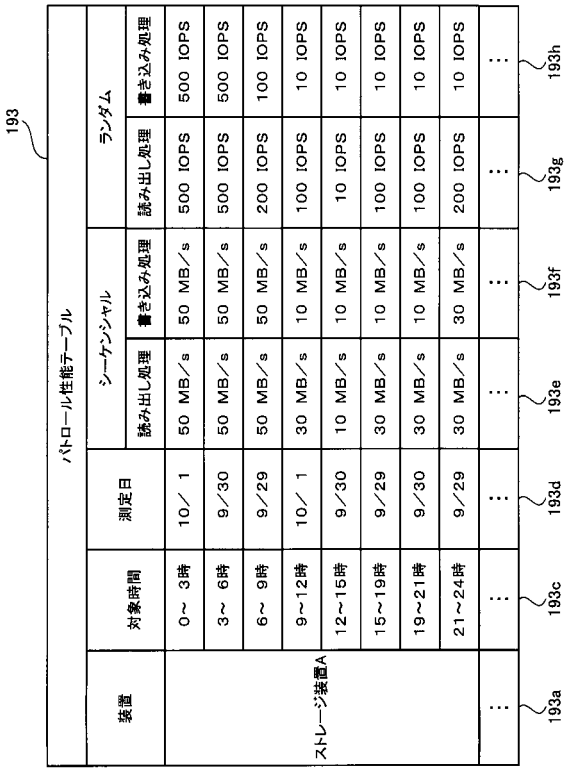
パトロール性能テーブル		シーケンシャル		ランダム	
装置		読み出し処理	書き込み処理	読み出し処理	書き込み処理
ストレージ装置A	191a	50 MB／s	50 MB／s	500 IOPS	500 IOPS
ストレージ装置B		30 MB／s	10 MB／s	100 IOPS	10 IOPS
…		…	…	…	…
		191e	191f	191g	191h

【図８】

192

パトロール性能テーブル			シーケンシャル		ランダム	
装置		領域	読み出し処理	書き込み処理	読み出し処理	書き込み処理
ストレージ装置A	192a	0～ 100[block]	50 MB／s	50 MB／s	500 IOPS	500 IOPS
		100～ 200[block]	30 MB／s	10 MB／s	100 IOPS	10 IOPS
		200～1000[block]	30 MB／s	10 MB／s	10 IOPS	10 IOPS
…		…	…	…	…	…
		192b	192e	192f	192g	192h

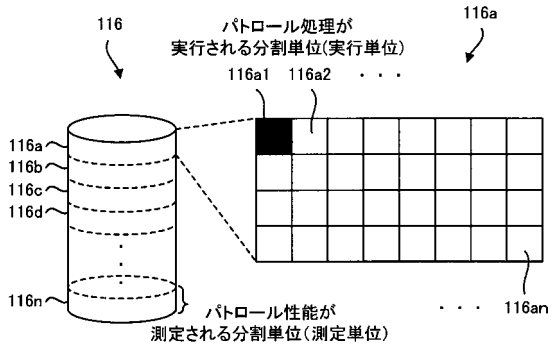
【図９】



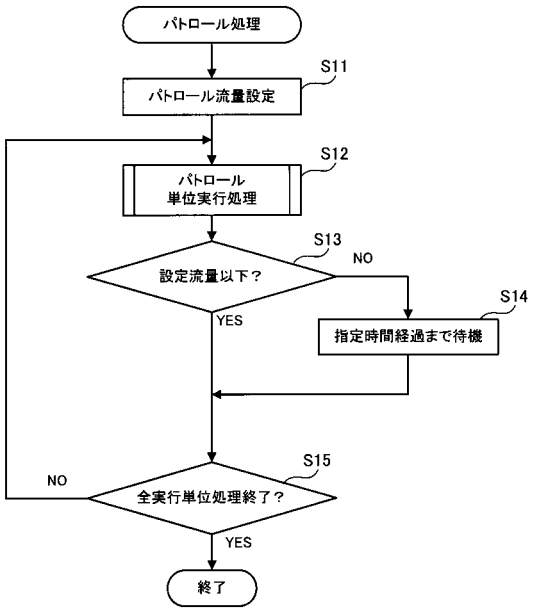
パトロール性能テーブル 193

装置	対象時間	測定日	シーケンシャル 読み出し処理	シーケンシャル 書き込み処理	ランダム 読み出し処理	ランダム 書き込み処理	
ストレージ装置A	0～ 3時	10／ 1	50 MB／s	50 MB／s	500 IOPS	500 IOPS	193h
	3～ 6時	9／30	50 MB／s	50 MB／s	500 IOPS	500 IOPS	193g
	6～ 9時	9／29	50 MB／s	50 MB／s	200 IOPS	100 IOPS	193f
	9～12時	10／ 1	30 MB／s	10 MB／s	100 IOPS	10 IOPS	193e
	12～15時	9／30	10 MB／s	10 MB／s	10 IOPS	10 IOPS	193d
	15～19時	9／29	30 MB／s	10 MB／s	100 IOPS	10 IOPS	193c
	19～21時	9／30	30 MB／s	10 MB／s	100 IOPS	10 IOPS	193a
	21～24時	9／29	30 MB／s	30 MB／s	200 IOPS	10 IOPS	

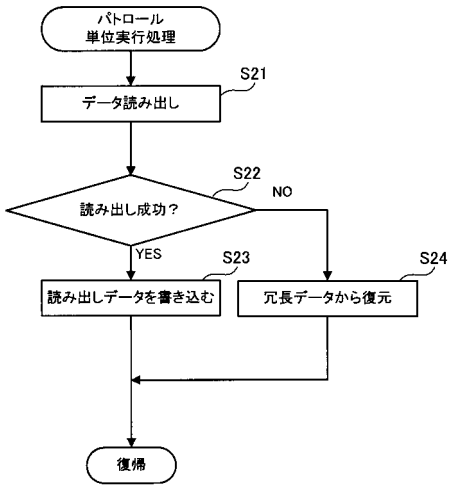
【図１０】



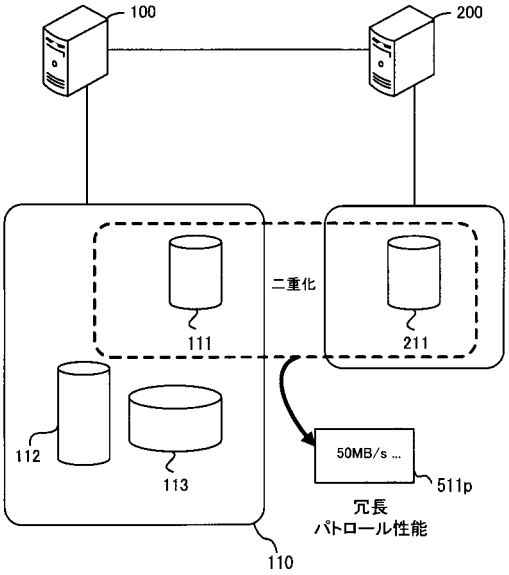
【図１１】



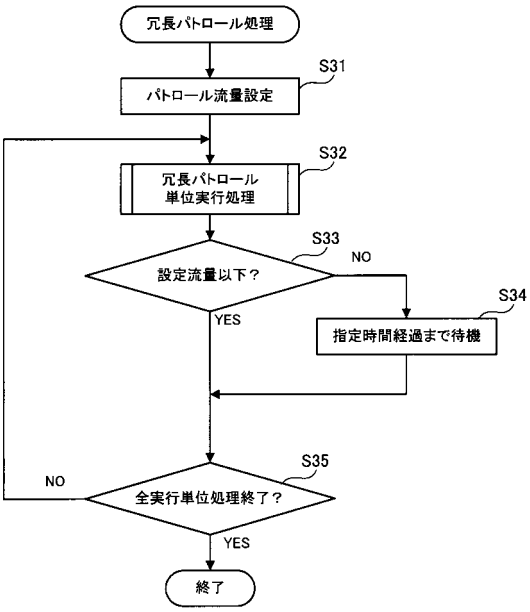
【図１２】



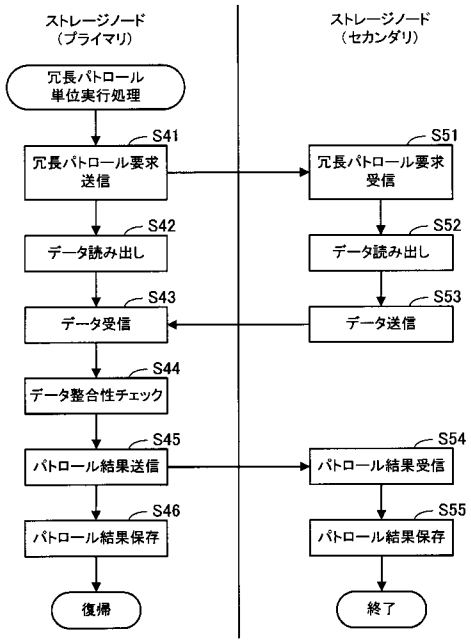
【図１３】


100
200
二重化
111
211
112
113
50MB/s ...
511p
冗長
パトロール性能
110

【図１４】


冗長パトロール処理
S31
パトロール流量設定
S32
冗長パトロール
単位実行処理
S33
設定流量以下？
NO
YES
S34
指定時間経過まで待機
S35
全実行単位処理終了？
NO
YES
終了

【図１５】


ストレージノード
（プライマリ）
ストレージノード
（セカンダリ）
冗長パトロール
単位実行処理
S41
冗長パトロール要求
送信
S51
冗長パトロール要求
受信
S42
データ読み出し
S52
データ読み出し
S43
データ受信
S53
データ送信
S44
データ整合性チェック
S45
パトロール結果送信
S54
パトロール結果受信
S46
パトロール結果保存
S55
パトロール結果保存
復帰
終了

フロントページの続き

(72)発明者　土屋　芳浩
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
(72)発明者　丸山　哲太郎
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
(72)発明者　武　理一郎
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

審査官　木村　貴俊

(56)参考文献　特開平０７－２１０３２５（ＪＰ，Ａ）
特開平０９－０６２４６１（ＪＰ，Ａ）
特開平０７－１０４９４７（ＪＰ，Ａ）
特開２００５－１５７９３３（ＪＰ，Ａ）
国際公開第２００４／１０４８４５（WO，Ａ１）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　３／０６－　３／０８
Ｇ０６Ｆ　１２／００－１２／１６
Ｇ０６Ｆ　１３／１０－１３／１４