(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5099128号
(P5099128)
(45)発行日　平成24年12月12日(2012.12.12)　(24)登録日　平成24年10月5日(2012.10.5)

(51)Int.Cl.　　F I
G06F 12/00 (2006.01)　　G06F 12/00 531D
G06F 3/06 (2006.01)　　G06F 3/06 301P
　　G06F 12/00 545A

請求項の数 18　(全 27 頁)

(21)出願番号　特願2009-512810 (P2009-512810)
(86)(22)出願日　平成19年4月20日(2007.4.20)
(86)国際出願番号　PCT/JP2007/058633
(87)国際公開番号　WO2008/136075
(87)国際公開日　平成20年11月13日(2008.11.13)
審査請求日　平成21年7月3日(2009.7.3)

前置審査

(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　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
(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　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
最終頁に続く

(54)【発明の名称】ストレージ管理プログラム、ストレージ管理装置およびストレージ管理方法

(57)【特許請求の範囲】
【請求項１】
同一内容の複数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ノードに分散して配置する分散ストレージシステムのデータ配置状況を管理するストレージ管理プログラムにおいて、
コンピュータを、
前記同一内容の複数のデータから、アクセス要求時にアクセス先として使用する主データとバックアップとして使用する副データとを示しており前記主データおよび前記副データそれぞれの配置先の前記ストレージノードを登録した管理情報を記憶する管理情報記憶手段、
前記ストレージノードの負荷情報を継続的に収集する負荷情報収集手段、
前記管理情報記憶手段が記憶する前記管理情報と前記負荷情報収集手段が収集した前記負荷情報とに基づいて、前記主データの配置先のストレージノードと前記副データの配置先のストレージノードとの間で各配置先のストレージノードの負荷の差が所定の許容量を超える、同一内容の前記主データと前記副データとの組を交換対象として検出する交換対象検出手段、
前記交換対象検出手段が検出した１組のデータの間で前記主データと前記副データとの役割を交換するように、前記管理情報記憶手段が記憶する前記管理情報を更新する管理情報更新手段、
前記複数のストレージノードのうちアクセス先のストレージノードを前記管理情報に基

づいて選択し当該アクセス先のストレージノードに格納されたアクセス先のデータが前記副データであるために当該アクセス先のストレージノードにより当該データへのアクセスが拒否されると前記管理情報の参照要求を送信し当該参照要求に応じて取得した前記管理情報に基づいてアクセス先のストレージノードを再選択するアクセスノードから、前記参照要求を受信すると、前記アクセスノードに更新後の前記管理情報を提供する手段、

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２】

前記交換対象検出手段は、前記ストレージノードのうち負荷最大ノードと負荷最小ノードとの間で負荷の差が前記所定の許容量を超えるか否か判断し、前記所定の許容量を超える場合、前記負荷最大ノードに配置された前記主データと前記負荷最小ノードに配置された前記副データとを交換対象の候補とすることを特徴とする請求項１記載のストレージ管理プログラム。

【請求項３】

前記交換対象検出手段は、交換対象の検出処理を所定の時間間隔で継続して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プログラム。

【請求項４】

前記管理情報では、データの特定に用いる論理ボリュームのアドレス空間が複数の論理セグメントに分割され、前記論理セグメント単位で前記同一内容の複数のデータが管理されており、

前記管理情報更新手段は、前記論理セグメント単位で前記主データと前記副データとの役割を交換する、

ことを特徴とする請求項１記載のストレージ管理プログラム。

【請求項５】

前記コンピュータを、更に、前記ストレージノードから当該ストレージノードに配置されたデータの格納位置の再編成を開始する通知を受け付ける通知受付手段として機能させ、

前記交換対象検出手段は、前記通知受付手段が通知を受け付けると、通知元の前記ストレージノードに配置された全ての前記主データを特定し、負荷の大きさに拘わらず、特定した前記主データと当該主データと同一内容の前記副データとの組を交換対象とする、

ことを特徴とする請求項１記載のストレージ管理プログラム。

【請求項６】

前記コンピュータを、更に、前記管理情報更新手段が前記管理情報を更新した後、役割が変更されたデータが配置されている前記ストレージノードに対して変更内容を通知する更新内容通知手段として機能させることを特徴とする請求項１記載のストレージ管理プログラム。

【請求項７】

同一内容の複数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ノードに分散して配置する分散ストレージシステムのデータ配置状況を管理するストレージ管理装置において、

前記同一内容の複数のデータから、アクセス要求時にアクセス先として使用する主データとバックアップとして使用する副データとを示しており前記主データおよび前記副データそれぞれの配置先の前記ストレージノードを登録した管理情報を記憶する管理情報記憶手段と、

前記ストレージノードの負荷情報を継続的に収集する負荷情報収集手段と、

前記管理情報記憶手段が記憶する前記管理情報と前記負荷情報収集手段が収集した前記負荷情報とに基づいて、前記主データの配置先のストレージノードと前記副データの配置先のストレージノードとの間で各配置先のストレージノードの負荷の差が所定の許容量を超える、同一内容の前記主データと前記副データとの組を交換対象として検出する交換対象検出手段と、

前記交換対象検出手段が検出した１組のデータの間で前記主データと前記副データとの

役割を交換するように、前記管理情報記憶手段が記憶する前記管理情報を更新する管理情報更新手段と、

前記複数のストレージノードのうちアクセス先のストレージノードを前記管理情報に基づいて選択し当該アクセス先のストレージノードに格納されたアクセス先のデータが前記副データであるために当該アクセス先のストレージノードにより当該データへのアクセスが拒否されると前記管理情報の参照要求を送信し当該参照要求に応じて取得した前記管理情報に基づいてアクセス先のストレージノードを再選択するアクセスノードから、前記参照要求を受信すると、前記アクセスノードに更新後の前記管理情報を提供する手段と、

を有することを特徴とするストレージ管理装置。

【請求項８】

前記交換対象検出手段は、前記ストレージノートのうち負荷最大ノードと負荷最小ノードとの間で負荷の差が前記所定の許容量を超えるか否か判断し、前記所定の許容量を超える場合、前記負荷最大ノードに配置された前記主データと前記負荷最小ノードに配置された前記副データとを交換対象の候補とすることを特徴とする請求項７記載のストレージ管理装置。

【請求項９】

前記交換対象検出手段は、交換対象の検出処理を所定の時間間隔で継続して実行することを特徴とする請求項７記載のストレージ管理装置。

【請求項１０】

前記管理情報では、データの特定に用いる論理ボリュームのアドレス空間が複数の論理セグメントに分割され、前記論理セグメント単位で前記同一内容の複数のデータが管理されており、

前記管理情報更新手段は、前記論理セグメント単位で前記主データと前記副データとの役割を交換する、

ことを特徴とする請求項７記載のストレージ管理装置。

【請求項１１】

前記ストレージノードから当該ストレージノードに配置されたデータの格納位置の再編成を開始する通知を受け付ける通知受付手段を更に有し、

前記交換対象検出手段は、前記通知受付手段が通知を受け付けると、通知元の前記ストレージノードに配置された全ての前記主データを特定し、負荷の大きさに拘わらず、特定した前記主データと当該主データと同一内容の前記副データとの組を交換対象とする、

ことを特徴とする請求項７記載のストレージ管理装置。

【請求項１２】

前記管理情報更新手段が前記管理情報を更新した後、役割が変更されたデータが配置されている前記ストレージノードに対して変更内容を通知する更新内容通知手段を更に有することを特徴とする請求項７記載のストレージ管理装置。

【請求項１３】

同一内容の複数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ノードに分散して配置する分散ストレージシステムのデータ配置状況を管理するコンピュータによるストレージ管理方法において、前記コンピュータが、

前記ストレージノードの負荷情報を継続的に収集し、

前記同一内容の複数のデータからアクセス要求時にアクセス先として使用する主データとバックアップとして使用する副データとを示しており前記主データおよび前記副データそれぞれの配置先の前記ストレージノードを登録した管理情報記憶手段が記憶する管理情報と、収集した前記負荷情報とに基づいて、前記主データの配置先のストレージノードと前記副データの配置先のストレージノードとの間で各配置先のストレージノードの負荷の差が所定の許容量を超える、同一内容の前記主データと前記副データとの組を交換対象として検出し、

検出した１組のデータの間で前記主データと前記副データとの役割を交換するように、前記管理情報記憶手段が記憶する前記管理情報を更新し、

前記複数のストレージノードのうちアクセス先のストレージノードを前記管理情報に基づいて選択し当該アクセス先のストレージノードに格納されたアクセス先のデータが前記副データであるために当該アクセス先のストレージノードにより当該データへのアクセスが拒否されると前記管理情報の参照要求を送信し当該参照要求に応じて取得した前記管理情報に基づいてアクセス先のストレージノードを再選択するアクセスノードから、前記参照要求を受信すると、前記アクセスノードに更新後の前記管理情報を提供する、

ことを特徴とするストレージ管理方法。

【請求項１４】

交換対象の検出処理では、前記ストレージノードのうち負荷最大ノードと負荷最小ノードとの間で負荷の差が前記所定の許容量を超えるか否か判断し、前記所定の許容量を超える場合、前記負荷最大ノードに配置された前記主データと前記負荷最小ノードに配置された前記副データとを交換対象の候補とすることを特徴とする請求項１３記載のストレージ管理方法。

【請求項１５】

交換対象の検出処理は所定の時間間隔で継続して実行することを特徴とする請求項１３記載のストレージ管理方法。

【請求項１６】

前記管理情報では、データの特定に用いる論理ボリュームのアドレス空間が複数の論理セグメントに分割され、前記論理セグメント単位で前記同一内容の複数のデータが管理されており、

前記管理情報の更新処理では、前記論理セグメント単位で前記主データと前記副データとの役割を交換する、

ことを特徴とする請求項１３記載のストレージ管理方法。

【請求項１７】

交換対象の検出処理では、前記ストレージノードから当該ストレージノードに配置されたデータの格納位置の再編成を開始する通知を受け付けていたときは、通知元の前記ストレージノードに配置された全ての前記主データを特定し、負荷の大きさに拘わらず、特定した前記主データと当該主データと同一内容の前記副データとの組を交換対象とすることを特徴とする請求項１３記載のストレージ管理方法。

【請求項１８】

前記管理情報の更新処理の後、役割が変更されたデータが配置されている前記ストレージノードに対して変更内容を通知することを特徴とする請求項１３記載のストレージ管理方法。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明はストレージシステムのデータ配置状況を管理するストレージ管理プログラム、ストレージ管理装置およびストレージ管理方法に関し、特に複数のストレージノードに分散してデータを配置する分散ストレージシステムのデータ配置状況を管理するストレージ管理プログラム、ストレージ管理装置およびストレージ管理方法に関する。

【背景技術】

【０００２】

現在、コンピュータによる情報処理が普及するに伴い、データを蓄積するストレージシステムの高機能化の要求が一層高まっている。例えば、大量のデータを読み書きする処理が高速であること、すなわち、高性能化が求められている。また、ストレージシステムを構成するハードウェアの一部に障害が発生しても蓄積したデータが消失しないこと、すなわち、高信頼化が求められている。

【０００３】

このような高性能化・高信頼化の要求に応えるストレージシステムとして、分散ストレージシステムが知られている。分散ストレージシステムでは、ネットワークで接続された

複数のストレージノードにデータが分散して配置される。また、分散ストレージシステムでは、同一内容のデータが複数のストレージノードに重複して配置されることが多い。すなわち、複数のストレージノードを用いて、負荷の分散化およびデータの冗長化が行われる。これにより、ストレージシステムの応答性能および信頼性を高めることができる。

【０００４】

ところで、分散ストレージシステムの運用時には、データの追加や削除に伴って、データの配置状況が絶えず変化する。ここで、頻繁にアクセス要求があるデータが特定のストレージノードに集中して配置されると、負荷の分散が不十分となり、応答性能が低下する。従って、分散ストレージシステムの応答性能を高く維持するためには、運用時に個々のストレージノードの負荷を考慮してデータの配置を管理する必要がある。

【０００５】

具体的には、分散ストレージシステムの応答性能を高く維持する管理方法として、以下の方法が知られている。第１の方法は、アクセス要求を受け付けたときに、動的にアクセス先のストレージノードを判定する方法である。すなわち、分散ストレージシステムは、アクセス要求があったデータが複数のストレージノードに重複して配置されている場合、アクセス要求時に最も負荷が小さいストレージノードをアクセス先とする（例えば、特許文献１，２参照）。一方、第２の方法は、ストレージノード間でアクセス要求の量ができる限り均等になるように、データ自体を自動的に移動する方法である（例えば、特許文献３参照）。これにより、特定のストレージノードに負荷が集中することを防止できる。

【特許文献１】特開昭６３－５６８７３号公報

【特許文献２】特開平１１－１６１５５５号公報

【特許文献３】特開２００５－２７６０１７号公報

【発明の開示】

【発明が解決しようとする課題】

【０００６】

しかし、上記特許文献１～３に記載の技術には以下のような問題がある。特許文献１，２に記載の動的な負荷分散の技術では、アクセス要求元とストレージノードとの間で、アクセス要求毎に必ずアクセス先を判定する処理が実行される。このため、データの配置管理を実現する仕組みが複雑になると共に、アクセス先を判定する処理だけ応答が遅くなる。また、特許文献３に記載のデータの自動移動の技術では、データの移動自体がストレージノードに負荷を与える。このため、データ移動中の応答が遅くなる。また、データの移動を繰り返すと、ディスク装置などのハードウェアへの負担が大きくなり、故障の原因ともなる。

【０００７】

本発明はこのような点に鑑みてなされたものであり、データの冗長化がなされている分散ストレージシステムにおいて、ストレージノード間の負荷分散を容易に実現するストレージ管理プログラム、ストレージ管理装置およびストレージ管理方法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】

【０００８】

本発明では、上記課題を解決するために、図１に示すストレージ管理プログラムが提供される。本発明に係るストレージ管理プログラムは、同一内容の複数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ノードに分散して配置する分散ストレージシステムのデータ配置状況を管理するものである。ストレージ管理プログラムを実行するコンピュータ１は、管理情報記憶手段１ａ、負荷情報収集手段１ｂ、交換対象検出手段１ｃ、管理情報更新手段１ｄおよび管理情報を提供する手段を有する。管理情報記憶手段１ａは、同一内容の複数のデータから、アクセス要求時にアクセス先として使用する主データとバックアップとして使用する副データとを示しており主データおよび副データそれぞれの配置先のストレージノードを登録した管理情報を記憶する。負荷情報収集手段１ｂは、ストレージノード２，３，４の負荷情報を継続的に収集する。交換対象検出手段１ｃは、管理情報

記憶手段１ａが記憶する管理情報と負荷情報収集手段１ｂが収集した負荷情報とに基づいて、主データの配置先のストレージノードと副データの配置先のストレージノードとの間で各配置先のストレージノードの負荷の差が所定の許容量を超える、同一内容の主データと副データとの組を交換対象として検出する。管理情報更新手段１ｄは、交換対象検出手段１ｃが検出した１組のデータの間で主データと副データとの役割を交換するように、管理情報記憶手段１ａが記憶する管理情報を更新する。管理情報を提供する手段は、複数のストレージノードのうちアクセス先のストレージノードを管理情報に基づいて選択し当該アクセス先のストレージノードに格納されたアクセス先のデータが副データであるために当該アクセス先のストレージノードにより当該データへのアクセスが拒否されると管理情報の参照要求を送信し当該参照要求に応じて取得した管理情報に基づいてアクセス先のストレージノードを再選択するアクセスノードから、参照要求を受信すると、アクセスノードに、更新後の管理情報を提供する。

【０００９】

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１によれば、負荷情報収集手段１ｂにより、ストレージノード２，３，４の負荷情報が継続的に収集される。交換対象検出手段１ｃにより、主データの配置先の負荷と副データの配置先の負荷との差が所定の許容量を超えるような、同一内容の主データと副データとの組が検出される。そして、管理情報更新手段１ｄにより、検出された１組のデータの間で主データと副データの役割が交換される。すなわち、交換前にバックアップ用であったデータがアクセス用のデータとなり、交換前にアクセス用であったデータがバックアップ用のデータとなる。更に、管理情報を提供する手段により、複数のストレージノードのうちアクセス先のストレージノードを管理情報に基づいて選択し当該アクセス先のストレージノードに格納されたアクセス先のデータが副データであるために当該アクセス先のストレージノードにより当該データへのアクセスが拒否されると管理情報の参照要求を送信し当該参照要求に応じて取得した管理情報に基づいてアクセス先のストレージノードを再選択するアクセスノードから、参照要求が受信されると、アクセスノードに、更新後の管理情報が提供される。

【００１０】

また、上記課題を解決するために、同一内容の複数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ノードに分散して配置する分散ストレージシステムのデータ配置状況を管理するストレージ管理装置において、同一内容の複数のデータから、アクセス要求時にアクセス先として使用する主データとバックアップとして使用する副データとを示しており主データおよび副データそれぞれの配置先のストレージノードを登録した管理情報を記憶する管理情報記憶手段と、ストレージノードの負荷情報を継続的に収集する負荷情報収集手段と、管理情報記憶手段が記憶する管理情報と負荷情報収集手段が収集した負荷情報とに基づいて、主データの配置先のストレージノードと副データの配置先のストレージノードとの間で各配置先のストレージノードの負荷の差が所定の許容量を超える、同一内容の主データと副データとの組を交換対象として検出する交換対象検出手段と、交換対象検出手段が検出した１組のデータの間で主データと副データとの役割を交換するように、管理情報記憶手段が記憶する管理情報を更新する管理情報更新手段と、複数のストレージノードのうちアクセス先のストレージノードを管理情報に基づいて選択し当該アクセス先のストレージノードに格納されたアクセス先のデータが副データであるために当該アクセス先のストレージノードにより当該データへのアクセスが拒否されると管理情報の参照要求を送信し当該参照要求に応じて取得した管理情報に基づいてアクセス先のストレージノードを再選択するアクセスノードから、参照要求を受信すると、アクセスノードに、更新後の管理情報を提供する手段と、を有することを特徴とするストレージ管理装置が提供される。

【００１１】

このようなストレージ管理装置によれば、負荷情報収集手段により、ストレージノードの負荷情報が継続的に収集される。交換対象検出手段により、主データの配置先のストレージノードの負荷と副データの配置先のストレージノードの負荷との差が所定の許容量を

超えるような、同一内容の主データと副データとの組が検出される。そして、管理情報更新手段により、検出された１組のデータの間で主データと副データの役割が交換される。すなわち、交換前にバックアップ用であったデータがアクセス用のデータとなり、交換前にアクセス用であったデータがバックアップ用のデータとなる。更に、複数のストレージノードのうちアクセス先のストレージノードを管理情報に基づいて選択し当該アクセス先のストレージノードに格納されたアクセス先のデータが副データであるために当該アクセス先のストレージノードにより当該データへのアクセスが拒否されると管理情報の参照要求を送信し当該参照要求に応じて取得した管理情報に基づいてアクセス先のストレージノードを再選択するアクセスノードから、参照要求が受信されると、アクセスノードに、更新後の管理情報が提供される。

【００１２】

また、上記課題を解決するために、同一内容の複数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ノードに分散して配置する分散ストレージシステムのデータ配置状況を管理するコンピュータによるストレージ管理方法において、コンピュータが、ストレージノードの負荷情報を継続的に収集し、同一内容の複数のデータからアクセス要求時にアクセス先として使用する主データとバックアップとして使用する副データとを示しており主データおよび副データそれぞれの配置先のストレージノードを登録した管理情報記憶手段が記憶する管理情報と、収集した負荷情報とに基づいて、主データの配置先のストレージノードと副データの配置先のストレージノードとの間で各配置先のストレージノードの負荷の差が所定の許容量を超える、同一内容の主データと副データとの組を交換対象として検出し、検出した１組のデータの間で主データと副データとの役割を交換するように、管理情報記憶手段が記憶する管理情報を更新し、複数のストレージノードのうちアクセス先のストレージノードを管理情報に基づいて選択し当該アクセス先のストレージノードに格納されたアクセス先のデータが副データであるために当該アクセス先のストレージノードにより当該データへのアクセスが拒否されると管理情報の参照要求を送信し当該参照要求に応じて取得した管理情報に基づいてアクセス先のストレージノードを再選択するアクセスノードから、参照要求を受信すると、アクセスノードに、更新後の管理情報を提供することを特徴とするストレージ管理方法が提供される。

【００１３】

このようなストレージ管理方法によれば、まずストレージノードの負荷情報が継続的に収集される。次に、主データの配置先のストレージノードの負荷と副データの配置先のストレージノードの負荷との差が所定の許容量を超えるような、同一内容の主データと副データとの組が検出される。そして、検出された１組のデータの間で主データと副データの役割が交換される。すなわち、交換前にバックアップ用であったデータがアクセス用のデータとなり、交換前にアクセス用であったデータがバックアップ用のデータとなる。更に、複数のストレージノードのうちアクセス先のストレージノードを管理情報に基づいて選択し当該アクセス先のストレージノードに格納されたアクセス先のデータが副データであるために当該アクセス先のストレージノードにより当該データへのアクセスが拒否されると管理情報の参照要求を送信し当該参照要求に応じて取得した管理情報に基づいてアクセス先のストレージノードを再選択するアクセスノードから、参照要求が受信されると、アクセスノードに、更新後の管理情報が提供される。

【発明の効果】

【００１４】

本発明では、アクセス用の主データが配置されたストレージノードとバックアップ用の副データが配置されたストレージノードとの間で負荷が所定の許容値を超えると、設定を変更し、主データと副データの役割を交換することとした。これにより、データを移動することなく、当該データのアクセス先のストレージノードが変更される。このとき、アクセス要求単位でアクセス先が判定されるわけではない。従って、負荷分散の処理自体が原因で応答が遅くなることを防止し、分散ストレージシステムの応答性能を確実に高めるこ

とができる。

【００１５】

本発明の上記および他の目的、特徴および利点は本発明の例として好ましい実施の形態を表す添付の図面と関連した以下の説明により明らかになるであろう。

【図面の簡単な説明】

【００１６】

【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】コントロールノードのハードウェア構成を示す図である。

【図５】論理ボリュームのデータ構造を示す第１の模式図である。

【図６】ストレージノードの機能を示すブロック図である。

【図７】コントロールノードおよびアクセスノードの機能を示すブロック図である。

【図８】スライス情報テーブルのデータ構造例を示す図である。

【図９】ボリューム情報テーブルのデータ構造例を示す図である。

【図１０】負荷情報テーブルのデータ構造例を示す図である。

【図１１】第１の種別変更処理の手順を示すフローチャートである。

【図１２】第２の種別変更処理の手順を示すフローチャートである。

【図１３】論理ボリュームのデータ構造を示す第２の模式図である。

【図１４】種別一斉変更処理の手順を示すフローチャートである。

【図１５】アクセスノードからのデータアクセスの流れを示すシーケンス図である。

【発明を実施するための最良の形態】

【００１７】

以下、本発明の実施の形態を図面を参照して説明する。まず、本実施の形態の概要について説明し、その後、本実施の形態の具体的な内容を説明する。

図１は、本実施の形態の概要を示す図である。図１に示す分散ストレージシステムは、コンピュータ１、ストレージノード２，３，４、コンピュータ５およびネットワーク６から構成される。コンピュータ１、ストレージノード２，３，４およびコンピュータ５は、ネットワーク６に接続されている。

【００１８】

コンピュータ１は、ストレージノード２，３，４に配置されたデータの配置状況を管理する。コンピュータ１は、管理情報記憶手段１ａ、負荷情報収集手段１ｂ、交換対象検出手段１ｃおよび管理情報更新手段１ｄを有する。

【００１９】

管理情報記憶手段１ａには、管理情報が格納されている。管理情報は、データの配置状況を管理する情報である。具体的には、管理情報には、データそれぞれの配置先のストレージノードが登録されている。また、同一内容の複数のデータがストレージノード２，３，４に分散されて配置されている場合、アクセス要求時にアクセス先として使用する主データとバックアップとして使用する副データとが指定されている。

【００２０】

負荷情報収集手段１ｂは、ストレージノード２，３，４の負荷情報を収集する。負荷情報には、例えば、ＣＰＵ（Central Processing Unit）の負荷、受け付けたアクセス要求の数、ネットワーク使用率などが含まれる。負荷情報収集手段１ｂは、最新の負荷情報を継続的に収集する。

【００２１】

交換対象検出手段１ｃは、管理情報記憶手段１ａに格納された管理情報と、負荷情報収集手段１ｂによって収集された負荷情報とに基づいて、所定の条件を具備する同一内容の主データと副データとの組を検出する。所定の条件とは、主データの配置先のストレージノードの負荷と、副データの配置先のストレージノードの負荷との差が、所定の許容量を超えることである。交換対象検出手段１ｃは、上記条件を具備する主データと副データと

の組が存在するか否かを調べる処理を継続的に実行する。

【００２２】

管理情報更新手段１ｄは、交換対象検出手段１ｃによって上記条件を具備する主データと副データとの組が検出されると、管理情報記憶手段１ａに格納された管理情報を更新する。具体的には、管理情報更新手段１ｄは、検出された１組のデータの間で、主データと副データとの役割を交換する。すなわち、交換前にバックアップ用であったデータがアクセス用のデータとなり、交換前にアクセス用であったデータがバックアップ用のデータとなる。

【００２３】

コンピュータ５は、コンピュータ１から管理情報を取得し、管理情報に基づいてストレージノード２，３，４に配置されたデータにアクセスする。例えば、利用しようとするデータの主データがストレージノード２に配置され副データがストレージノード３に配置されているとき、コンピュータ５はストレージノード２にアクセスする。アクセスによって主データが更新されたときは、ストレージノード間で通信が行われ、更新内容が自動的に副データに反映される。ただし、コンピュータ５が主データと副データとを同時に更新するようにしてもよい。

【００２４】

ここで、データ＃１の主データがストレージノード２（管理情報では“Ａ”と表記）に副データがストレージ３（管理情報では“Ｂ”と表記）に配置され、データ＃２の主データがストレージノード３に副データがストレージノード４（管理情報では“Ｃ”と表記）に配置され、データ＃３の主データがストレージノード２に副データがストレージノード４に配置されているとする。また、ストレージノード２，３，４のうち、ストレージノード２の負荷が最も高くストレージノード４の負荷が最も低く、両者の負荷の差が所定の許容量を超えているとする。

【００２５】

このとき、コンピュータ１は、管理情報を更新して、データ＃３の主データと副データとの役割を交換する。すなわち、交換後は、データ＃３の主データがストレージノード４に副データがストレージノード２に配置されることになる。これによって、コンピュータ５はデータ＃３を利用するとき、ストレージノード４にアクセスすることになる。

【００２６】

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１によれば、負荷情報収集手段１ｂにより、ストレージノード２，３，４の負荷情報が継続的に収集される。交換対象検出手段１ｃにより、主データの配置先の負荷と副データの配置先の負荷との差が所定の許容量を超えるような、同一内容の主データと副データとの組が検出される。そして、管理情報更新手段１ｄにより、検出された１組のデータの間で主データと副データの役割が交換される。すなわち、交換前にバックアップ用であったデータがアクセス用のデータとなり、交換前にアクセス用であったデータがバックアップ用のデータとなる。

【００２７】

これにより、データを移動することなく、当該データのアクセス先のストレージノードが変更される。このとき、アクセス要求単位でアクセス先が判定されるわけではない。従って、負荷分散の処理自体が原因で応答が遅くなることを防止し、分散ストレージシステムの応答性能を確実に高めることができる。

【００２８】

以下、本実施の形態を図面を参照して詳細に説明する。

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に示す分散ストレージシステムは、同一内容の複数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ノードに分散して配置することで、信頼性と処理性能とを向上させたストレージシステムである。

【００２９】

本実施の形態に係る分散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３

０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。

【００３０】

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。

【００３１】

ストレージ装置１１０には、複数の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、複数のＨＤＤ２１１，２１２，２１３，２１４が実装されている。ストレージ装置３１０には、複数のＨＤＤ３１１，３１２，３１３，３１４が実装されている。ストレージ装置４１０には、複数のＨＤＤ４１１，４１２，４１３，４１４が実装されている。各ストレージ装置１１０，２１０，３１０，４１０は、内蔵する複数のＨＤＤを用いたＲＡＩＤ（Redundant Array of Independent Disks）システムである。本実施の形態では、各ストレージ装置１１０，２１０，３１０，４１０は、ＲＡＩＤ５のディスク管理サービスを提供する。

【００３２】

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。論理ボリュームについては、後で詳細に説明する。

【００３３】

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。

【００３４】

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、運用時に必要な各種設定を行うことができる。

【００３５】

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、管理ノード３０および端末装置２１，２２，２３のハードウェア構成について説明する。

【００３６】

図３は、ストレージノードのハードウェア構成を示す図である。ストレージノード１００は、ＣＰＵ１０１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。

【００３７】
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。
【００３８】
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～ＲＡＩＤ５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。
【００３９】
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。
【００４０】
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。
【００４１】
ストレージノード２００，３００，４００も、ストレージノード１００と同様のハードウェア構成によって実現できる。
図４は、コントロールノードのハードウェア構成を示す図である。コントロールノード５００は、ＣＰＵ５０１によって装置全体が制御されている。ＣＰＵ５０１には、バス５０７を介してＲＡＭ５０２、ＨＤＤ５０３、グラフィック処理装置５０４、入力インタフェース５０５および通信インタフェース５０６が接続されている。
【００４２】
ＲＡＭ５０２には、ＣＰＵ５０１に実行させるＯＳ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、ＲＡＭ５０２には、ＣＰＵ５０１による処理に必要な各種データが格納される。ＨＤＤ５０３には、ＯＳやアプリケーションのプログラムが格納される。
【００４３】
グラフィック処理装置５０４には、モニタ５１が接続されている。グラフィック処理装置５０４は、ＣＰＵ５０１からの命令に従って、画像をモニタ５１の画面に表示させる。入力インタフェース５０５には、キーボード５２とマウス５３とが接続されている。入力インタフェース５０５は、キーボード５２やマウス５３から送られてくる信号を、バス５０７を介してＣＰＵ５０１に送信する。通信インタフェース５０６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５０６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。
【００４４】
アクセスノード６００、管理ノード３０および端末装置２１，２２，２３も、コントロールノード５００と同様のハードウェア構成によって実現できる。ただし、アクセスノード６００は、ネットワーク１０と接続するための通信インタフェースに加えて、ネットワーク２０と接続するための通信インタフェースを更に備えている。
【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。
ここで、コントロールノード５００がアクセスノード６００に対して提供する論理ボリュームについて説明する。論理ボリュームは、ストレージノード１００，２００，３００

，４００によって分散管理されているデータを、アクセスノード６００から容易に利用できるようにするための仮想的なボリュームである。

【００４６】

図５は、論理ボリュームのデータ構造を示す第１の模式図である。論理ボリューム７００には、“ＶＶ－Ａ”という論理ボリュームＩＤが付与されている。また、ストレージノード１００，２００，３００，４００には、それぞれ“ＳＮ－Ａ”、“ＳＮ－Ｂ”、“ＳＮ－Ｃ”、“ＳＮ－Ｄ”というノードＩＤが付与されている。

【００４７】

各ストレージノード１００，２００，３００，４００に接続されたストレージ装置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６つのスライスに分割されて、個々のストレージノード内で管理されている。

【００４８】

図５の例では、ストレージ装置１１０内の記憶領域は、６つのスライス１２１～１２６に分割されている。ストレージ装置２１０内の記憶領域は、６つのスライス２２１～２２６に分割されている。ストレージ装置３１０内の記憶領域は、６つのスライス３２１～３２６に分割されている。ストレージ装置４１０内の記憶領域は、６つのスライス４２１～４２６に分割されている。

【００４９】

論理ボリューム７００は、セグメント７１０，７２０，７３０，７４０，７５０，７６０という単位で構成される。セグメント７１０，７２０，７３０，７４０，７５０，７６０は、それぞれプライマリスライス７１１，７２１，７３１，７４１，７５１，７６１とセカンダリスライス７１２，７２２，７３２，７４２，７５２，７６２との組で構成される。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ように配置を行う。

【００５０】

図５の例では、スライスＩＤを、“Ｐ”または“Ｓ”のアルファベットと数字との組み合わせで示している。“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。

【００５１】

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１１０のスライス１２１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３１０のスライス３２２に対応付けられている。

【００５２】

そして、ストレージ装置１１０，２１０，３１０，４１０には、自己のスライスに対応付けられた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が格納される。

次に、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のモジュール構成について説明する。

【００５３】

図６は、ストレージノードの機能を示すブロック図である。ストレージノード１００は、スライス情報記憶部１３０、データアクセス部１４０、スライス情報管理部１５０および負荷監視部１６０を有する。

【００５４】

スライス情報記憶部１３０には、ストレージ装置１１０が有するスライスについてのスライス情報が格納される。スライス情報には、スライスを特定するためのアドレスやスライスへの割り当ての種別（プライマリスライスまたはセカンダリスライスのいずれか）な

どの情報が含まれている。また、プライマリスライスについては、対応するセカンダリスライスを管理しているストレージノードに関する情報も含まれている。

【００５５】

データアクセス部１４０は、アクセスノード６００からのアクセスを受け付けると、スライス情報記憶部１３０に格納されたスライス情報を参照して、ストレージ装置１１０に格納されたデータを操作する。

【００５６】

具体的には、データアクセス部１４０は、アクセスノード６００からアドレスを指定したデータの読み込み要求（Ｒｅａｄ要求）を受け付けると、指定されたアドレスが属するスライスがプライマリスライスであるか否か判断する。プライマリスライスである場合には、データアクセス部１４０は、指定されたアドレスに対応するデータをストレージ装置１１０から取得し、アクセスノード６００に送信する。プライマリスライスでない場合には、データアクセス部１４０は、アドレスの指定が不適切である旨をアクセスノード６００に通知する。

【００５７】

また、データアクセス部１４０は、アクセスノード６００からアドレスおよび書き込み内容を指定したデータの書き込み要求（Ｗｒｉｔｅ要求）を受け付けると、ストレージ装置１１０内の指定されたアドレスに対応する位置にデータの書き込みを試みる。そして、データアクセス部１４０は、書き込みの結果を、アクセスノード６００に通知する。ここで、指定されたアドレスが属するスライスがプライマリスライスである場合には、対応するセカンダリスライスを管理するストレージノードに対して、同様の書き込みを行うよう指示する。これによって、プライマリスライスとセカンダリスライスとの内容が一致するように維持される。

【００５８】

スライス情報管理部１５０は、ストレージノード１００の稼働状態を定期的にコントロールノード５００に通知する。また、スライス情報管理部１５０は、コントロールノード５００からスライス情報の取得要求があると、スライス情報記憶部１３０に格納されたスライス情報を送信する。また、スライス情報管理部１５０は、スライス情報の更新の指示を受け付けると、指示された更新内容を、スライス情報記憶部１３０に格納されたスライス情報に反映させる。

【００５９】

また、スライス情報管理部１５０は、ストレージ装置１１０がＨＤＤへのアクセスを効率化するためにデータ格納位置の再編成を開始すると、その旨をコントロールノード５００に通知する。再編成が終了すると、再編成後の各スライスのアドレスを、スライス情報記憶部１３０に格納されたスライス情報に反映させると共に、更新後のスライス情報をコントロールノード５００に対して送信する。

【００６０】

負荷監視部１６０は、ストレージノード１００の負荷を継続的に監視する。例えば、負荷監視部１６０は、ＣＰＵ１０１の稼働率、ＨＤＤインタフェース１０３に対するＩ／Ｏ（Input / Output）要求の回数、通信インタフェース１０６の使用率などを監視する。そして、負荷監視部１６０は、ストレージノード１００の負荷を示す負荷情報を、定期的にコントロールノード５００に対して送信する。なお、スライス情報管理部１５０による稼働状態の通知と負荷監視部１６０による負荷情報の送信とを兼ねるように、コントロールノード５００に対して通信を行うことも可能である。

【００６１】

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。

図７は、コントロールノードおよびアクセスノードの機能を示すブロック図である。

【００６２】

コントロールノード５００は、論理ボリューム記憶部５１０、負荷情報記憶部５２０、

論理ボリューム管理部５３０、負荷情報収集部５４０および種別変更部５５０を有する。

論理ボリューム記憶部５１０には、１つ以上の論理ボリュームが格納される。論理ボリュームでは、ストレージ装置１１０，２１０，３１０，４１０が管理する記憶領域を一元的に扱うために、仮想的なアドレスである論理アドレスによって各セグメントが管理されている。論理ボリュームには、セグメントを特定するための論理アドレスやセグメントに属するプライマリスライスおよびセカンダリスライスを特定するための情報などが含まれている。

【００６３】

負荷情報記憶部５２０には、ストレージノード１００，２００，３００，４００それぞれの最新の負荷を示す負荷情報が格納される。負荷情報記憶部５２０に格納される負荷情報は、ストレージノード１００，２００，３００，４００から取得したものである。

【００６４】

論理ボリューム管理部５３０は、ストレージノード１００，２００，３００，４００から稼働状態を示す通知をネットワーク１０経由で受信する。これにより、論理ボリューム管理部５３０は、ストレージノード１００，２００，３００，４００が正常に稼働しているか否かを知る。また、論理ボリューム管理部５３０は、必要に応じてストレージノード１００，２００，３００，４００からスライス情報を取得し、論理ボリューム記憶部５１０に格納された論理ボリュームを更新する。また、論理ボリューム管理部５３０は、論理ボリューム記憶部５１０が更新されると、更新によって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。

【００６５】

また、論理ボリューム管理部５３０は、ストレージノード１００，２００，３００，４００から再編成の開始の通知を受け付けると、負荷情報記憶部５２０に再編成中であることを示す情報を格納する。これは、再編成が開始されると、そのストレージ装置の負荷が高くなり、アクセスに対する応答性能が極端に低下することを考慮したものである。そして、論理ボリューム管理部５３０は、再編成後のスライス情報をストレージノード１００，２００，３００，４００から受信すると、受信したスライス情報を論理ボリューム記憶部５１０に反映させると共に、負荷情報記憶部５２０から再編成中であることを示す情報を削除する。

【００６６】

また、論理ボリューム管理部５３０は、アクセスノード６００から論理ボリュームの参照要求を受け付けると、論理ボリューム記憶部５１０に格納された論理ボリュームを、アクセスノード６００に対して送信する。

【００６７】

負荷情報収集部５４０は、ストレージノード１００，２００，３００，４００が定期的に送信する負荷情報を、ネットワーク１０経由で受信する。そして、負荷情報収集部５４０は、受信した負荷情報を負荷情報記憶部５２０に格納する。

【００６８】

種別変更部５５０は、論理ボリューム記憶部５１０に格納された論理ボリュームと負荷情報記憶部５２０に格納された負荷情報とを継続的に監視し、プライマリスライスとセカンダリスライスとを交換すべきセグメントを判定する。具体的には、種別変更部５５０は、プライマリスライスを割り当てたストレージノードの負荷とセカンダリスライスを割り当てたストレージノードの負荷との差が所定の許容量を超える場合に、そのプライマリスライスとセカンダリスライスとを交換すべきと判定する。そして、種別変更部５５０は、判定結果に応じて論理ボリューム記憶部５１０に格納された論理ボリュームを更新する。判定方法の詳細については、後で詳細に説明する。

【００６９】

なお、論理ボリューム管理部５３０および負荷情報収集部５４０は、ストレージノード１００，２００，３００，４００から各種情報が送られてくるのを待つ代わりに、定期的にストレージノード１００，２００，３００，４００にアクセスして各種情報を収集する

ようにしてもよい。

【００７０】

アクセスノード６００は、論理ボリューム記憶部６１０およびデータアクセス制御部６２０を有する。

論理ボリューム記憶部６１０には、コントロールノード５００の論理ボリューム記憶部５１０に格納されている論理ボリュームと同じ情報が格納される。

【００７１】

データアクセス制御部６２０は、実行中のプログラムからデータへのアクセス要求があると、まず論理ボリューム記憶部６１０に論理ボリュームが格納されているか否か確認する。論理ボリュームが格納されていない場合、データアクセス制御部６２０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を論理ボリューム記憶部６１０に格納する。

【００７２】

そして、データアクセス制御部６２０は、論理ボリュームに基づいてアクセス先のストレージノードを特定する。すなわち、利用するデータが属するセグメントを特定し、特定したセグメントのプライマリスライスを管理するストレージノードを特定する。その後、データアクセス制御部６２０は、特定したストレージノードにアクセスする。ここで、アクセスに失敗した場合、コントロールノード５００から論理ボリュームを取得した後にデータの配置状況が変化したことが考えられるため、データアクセス制御部６２０は、コントロールノード５００から最新の論理ボリュームを取得して、再びストレージノードへのアクセスを試みる。

【００７３】

図８は、スライス情報テーブルのデータ構造例を示す図である。図８に示すスライス情報テーブル１３１は、ストレージノード１００のスライス情報記憶部１３０に格納されている。スライス情報テーブル１３１には、ディスクを示す項目、物理アドレスを示す項目、ブロック数を示す項目、ボリュームを示す項目、論理アドレスを示す項目、種別を示す項目およびセカンダリを示す項目が設けられている。各項目の横方向に並べられた情報同士が互いに関連付けられて、１つのスライスについてのスライス情報を構成する。

【００７４】

ディスクを示す項目には、ＨＤＤを識別するディスクＩＤが設定される。物理アドレスを示す項目には、スライスの先頭ブロックを示す物理的なアドレスが設定される。ブロック数を示す項目には、スライスに含まれるブロックの数が設定される。

【００７５】

ボリュームを示す項目には、スライスに対応付けられたセグメントが属する論理ボリュームの論理ボリュームＩＤが設定される。論理アドレスを示す項目には、スライスに対応付けられたセグメントの先頭の論理アドレスが設定される。種別を示す項目には、“Ｐ”または“Ｓ”のいずれかの値が設定される。“Ｐ”はプライマリスライスを意味し、“Ｓ”はセカンダリスライスを意味する。

【００７６】

セカンダリを示す項目には、種別がプライマリスライスである場合、対応するセカンダリスライスの割り当て先の情報が設定される。具体的には、ストレージノードのノードＩＤ、ＨＤＤのディスクＩＤ、スライスの先頭ブロックを示す物理アドレスが設定される。種別がセカンダリスライスである場合、セカンダリを示す項目は空欄となる。

【００７７】

スライス情報テーブル１３１に格納されるスライス情報は、スライス情報管理部１５０によって適宜更新される。例えば、ディスクが“ｓｄ－ａ”、物理アドレスが“３０７２”、ブロック数が“５１２”、ボリュームが“ＶＶ－１”、論理アドレスが“４０９６”、種別が“Ｓ”という情報が格納される。これは、ディスクＩＤ“ｓｄ－ａ”のディスクの３０７２ブロックから３５８３ブロックまでの記憶領域が１つのスライスを構成し、そのスライスに、論理アドレスが４０９６ブロックから４６０７ブロックまでのセグメント

がセカンダリスライスとして割り当てられていることを示している。

【００７８】

図９は、ボリューム情報テーブルのデータ構造例を示す図である。図９に示す論理ボリュームテーブル５１１は、論理ボリュームＩＤが“ＶＶ－１”の論理ボリュームについてのテーブルである。論理ボリュームテーブル５１１は、コントロールノード５００の論理ボリューム記憶部５１０に格納されている。論理ボリュームテーブル５１１には、セグメントを示す項目、論理アドレスを示す項目、ブロック数を示す項目、種別を示す項目、ノードを示す項目、ディスクを示す項目および物理アドレスを示す項目が設けられている。各項目の横方向に並べられた情報同士が互いに関連付けられている。

【００７９】

セグメントを示す項目には、セグメントを識別するセグメントＩＤが設定される。論理アドレスを示す項目には、セグメントの先頭の論理アドレスが設定される。ブロック数を示す項目には、セグメントに含まれるブロックの数が設定される。

【００８０】

種別を示す項目には、“Ｐ”または“Ｓ”のいずれかの値が設定される。ノードを示す項目には、割り当て先のストレージノードを識別するノードＩＤが設定される。ディスクを示す項目には、ストレージノード内でＨＤＤを識別するディスクＩＤが設定される。物理アドレスを示す項目には、割り当て先のスライスの先頭ブロックを示す物理的なアドレスが設定される。

【００８１】

論理ボリュームテーブル５１１に格納される情報は、ストレージノード１００，２００，３００，４００から取得したスライス情報に基づいて、論理ボリューム管理部５３０によって生成される。

【００８２】

図１０は、負荷情報テーブルのデータ構造例を示す図である。図１０に示す負荷情報テーブル５２１は、コントロールノード５００の負荷情報記憶部５２０に格納されている。負荷情報テーブル５２１には、ノードを示す項目、ＣＰＵ負荷を示す項目、Ｉ／Ｏ数を示す項目、期間Ｉ／Ｏ数を示す項目およびＩ／Ｆ使用率を示す項目が設けられている。各項目の横方向に並べられた情報同士が互いに関連付けられて、１つのストレージノードについての負荷情報を構成する。

【００８３】

ノードを示す項目には、ストレージノードを識別するノードＩＤが設定される。ＣＰＵ負荷を示す項目には、ＣＰＵの現在の稼働率が設定される。Ｉ／Ｏ数を示す項目には、過去に発生したＩ／Ｏの総回数が設定される。期間Ｉ／Ｏ数を示す項目には、所定期間内に発生したＩ／Ｏの回数が設定される。例えば、直近２４時間に発生したＩ／Ｏの回数が設定される。Ｉ／Ｆ使用率を示す項目には、ネットワークの現在の使用率が設定される。

【００８４】

負荷情報テーブル５２１に格納される負荷情報は、負荷情報収集部５４０によって適宜更新される。例えば、ノードが“ＳＮ－Ａ”、ＣＰＵ負荷が“７５％”、Ｉ／Ｏ数が“１００００”、期間Ｉ／Ｏ数が“１０００”、Ｉ／Ｆ使用率が“６６％”という情報が格納される。

【００８５】

次に、以上のような構成およびデータ構造のシステムにおいて実行される処理の詳細を説明する。最初に、コントロールノード５００の種別変更部５５０が、ストレージノードの負荷に応じてプライマリスライスとセカンダリスライスとを交換する処理について説明する。

【００８６】

図１１は、第１の種別変更処理の手順を示すフローチャートである。この種別変更処理は、種別変更部５５０によって定期的に実行される。以下、図１１に示す処理をステップ番号に沿って説明する。

【００８７】
［ステップＳ１１］種別変更部５５０は、負荷情報記憶部５２０に格納された負荷情報テーブル５２１を参照して、負荷が最も高いストレージノードを特定する。
［ステップＳ１２］種別変更部５５０は、負荷情報テーブル５２１を参照して、負荷が最も低いストレージノードを特定する。
【００８８】
［ステップＳ１３］種別変更部５５０は、ステップＳ１１で特定した最高負荷のストレージノードの負荷と、ステップＳ１２で特定した最低負荷のストレージノードの負荷との差が、所定の閾値を超えるか否か判断する。閾値を超える場合には、処理がステップＳ１４に進められる。閾値以下である場合には、処理が終了する。
【００８９】
［ステップＳ１４］種別変更部５５０は、論理ボリューム記憶部５１０に格納された論理ボリュームテーブル５１１を参照して、ステップＳ１１で特定した最高負荷のストレージノードにプライマリスライスが割り当てられ、ステップＳ１２で特定した最低負荷のストレージノードのセカンダリスライスが割り当てられている、セグメントが存在するか否か判断する。存在する場合には、処理がステップＳ１５に進められる。存在しない場合には、処理が終了する。
【００９０】
［ステップＳ１５］種別変更部５５０は、論理ボリュームテーブル５１１を更新して、ステップＳ１４の判断条件を具備するプライマリスライスとセカンダリスライスとの間で種別を交換する。
【００９１】
［ステップＳ１６］論理ボリューム管理部５３０は、論理ボリュームテーブル５１１が更新されたことを検知し、ステップＳ１５で交換されたプライマリスライスおよびセカンダリスライスそれぞれを管理する２つのストレージノードに対して更新内容を通知する。
【００９２】
ここで、負荷を比較する際に、ＣＰＵ負荷、Ｉ／Ｏ数、期間Ｉ／Ｏ数およびＩ／Ｆ使用率のいずれの指標を用いるかは、分散ストレージシステムの管理者によって予め種別変更部５５０に設定されている。管理者は、いずれか１つの指標を用いて比較を行うように設定してもよいし、２つ以上の指標から所定の計算式によって計算される値を用いて比較を行うように設定してもよい。
【００９３】
なお、図１１に示した方法では、プライマリスライスを最高負荷のストレージノードから選択し、セカンダリスライスを最低負荷のストレージノードから選択することとしたが、セカンダリスライスを、最低負荷のストレージノード以外のストレージノードからも選択できるようにすることも考えられる。すなわち、プライマリスライスを最高負荷のストレージノードから選択し、セカンダリスライスを最高負荷のストレージノードとの負荷の差が許容量を超えている全てのストレージノードから選択できるようにしてもよい。このとき、複数のセカンダリスライスの候補がある場合には、最高負荷のストレージノードとの負荷の差ができる限り大きいストレージノードから選択するようにすればよい。
【００９４】
このようにして、種別変更部５５０は、最高負荷のストレージノードと最低負荷のストレージノードとの負荷の差が許容量を超えており、最高負荷のストレージノードにプライマリスライスが割り当てられ最低負荷のストレージノードにセカンダリスライスが割り当てられているセグメントが存在する場合には、両者の種別を交換する。これにより、最高負荷のストレージノードに対するアクセスの少なくとも一部が、最低負荷のストレージノードに振り向けられる。このとき、プライマリスライスとセカンダリスライスの内容は同一であるため、論理ボリュームを更新するのみでよく、ストレージ装置１１０，２１０，３１０，４１０に格納されたデータ自体を移動する必要はない。
【００９５】

交換対象となるプライマリスライスとセカンダリスライスとを検出する方法は、上記の方法以外にも、さまざまな方法が考えられる。次に、他の検出方法の例を示す。

図１２は、第２の種別変更処理の手順を示すフローチャートである。この種別変更処理は、種別変更部５５０によって定期的に実行される。以下、図１２に示す処理をステップ番号に沿って説明する。

【００９６】

［ステップＳ２１］種別変更部５５０は、論理ボリューム記憶部５１０に格納された論理ボリュームテーブル５１１を参照して、ストレージノードを１つ選択する。

［ステップＳ２２］種別変更部５５０は、論理ボリュームを参照して、ステップＳ２１で選択したストレージノードにプライマリスライスまたはセカンダリスライスを割り当てているセグメントを１つ選択する。

【００９７】

［ステップＳ２３］種別変更部５５０は、負荷情報記憶部５２０に格納された負荷情報テーブル５２１を参照して、ステップＳ２２で選択したセグメントのプライマリスライスの割り当て先のストレージノードとセカンダリスライスの割り当て先のストレージノードとの間で負荷の差を計算する。なお、セカンダリスライスの割り当て先のストレージノードの方が負荷が高い場合には、負荷の差は０とする。

【００９８】

［ステップＳ２４］種別変更部５５０は、ステップＳ２２で条件を具備する全てのセグメントを選択したか否か判断する。全てのセグメントを選択した場合には、処理がステップＳ２５に進められる。未選択のセグメントが存在する場合には、処理がステップＳ２２に進められる。

【００９９】

［ステップＳ２５］種別変更部５５０は、ステップＳ２３で計算した負荷の差の最大値が、所定の閾値を超えるか否か判断する。閾値を超える場合には、処理がステップＳ２６に進められる。閾値以下である場合には、処理がステップＳ２８に進められる。

【０１００】

［ステップＳ２６］種別変更部５５０は、論理ボリュームテーブル５１１を更新して、ステップＳ２３で計算した負荷の差が最大となったセグメントのプライマリスライスとセカンダリスライスとの間で種別を交換する。

【０１０１】

［ステップＳ２７］論理ボリューム管理部５３０は、論理ボリュームテーブル５１１が更新されたことを検知し、ステップＳ２６で交換されたプライマリスライスおよびセカンダリスライスそれぞれを管理する２つのストレージノードに対して更新内容を通知する。

【０１０２】

［ステップＳ２８］種別変更部５５０は、ステップＳ２１で全てのストレージノードを選択したか否か判断する。全てのストレージノードを選択した場合には、処理が終了する。未選択のストレージノードが存在する場合には、処理がステップＳ２９に進められる。

【０１０３】

［ステップＳ２９］種別変更部５５０は、一定時間待機する。その後、処理がステップＳ２１に進められる。

なお、上記ステップＳ２１では、ストレージノードを選択する順序を、負荷が高い順または負荷が低い順とすることも考えられる。

【０１０４】

このようにして、種別変更部５５０は、ストレージノードを順次確認し、プライマリスライスが割り当てられているストレージノードとセカンダリスライスが割り当てられているストレージノードとの負荷の差が許容量を超えているセグメントが存在する場合には、両者の種別を交換する。これにより、最高負荷のストレージノードにプライマリスライスが割り当てられ最低負荷のストレージノードにセカンダリスライスが割り当てられているセグメントが存在しない場合でも、負荷を分散させることができる。

【０１０５】
図１３は、論理ボリュームのデータ構造を示す第２の模式図である。図１３に示すセグメントの割り当て状況は、上記の種別変更処理が実行されることで、図５に示したセグメントの割り当て状況が変化したものである。図１３の例では、セグメント７５０のプライマリスライスとセカンダリスライスとが交換されている。すなわち、交換前は、プライマリスライス７５１がストレージ装置１１０に割り当てられ、セカンダリスライス７５２がストレージ装置４１０に割り当てられていたが、交換後は、プライマリスライス７５１がストレージ装置４１０に割り当てられ、セカンダリスライス７５２がストレージ装置１１０に割り当てられている。これにより、ストレージノード１００の負荷が軽減される。
【０１０６】
次に、再編成中のストレージ装置にプライマリスライスを割り当てないようにして、応答性能の低下を避ける処理について説明する。
図１４は、種別一斉変更処理の手順を示すフローチャートである。以下、図１４に示す処理をステップ番号に沿って説明する。
【０１０７】
［ステップＳ３１］論理ボリューム管理部５３０は、ストレージノード１００，２００，３００，４００のいずれかから再編成の開始を示す通知を受け付けると、負荷情報記憶部５２０に再編成の開始を示す情報を格納する。
【０１０８】
［ステップＳ３２］種別変更部５５０は、論理ボリューム記憶部５１０に格納された論理ボリュームテーブル５１１を参照して、再編成中のストレージノードにプライマリスライスを割り当てているセグメントを１つ選択する。
【０１０９】
［ステップＳ３３］種別変更部５５０は、論理ボリュームテーブル５１１を更新して、ステップＳ３２で選択したセグメントのプライマリスライスとセカンダリスライスとの間で種別を交換する。
【０１１０】
［ステップＳ３４］論理ボリューム管理部５３０は、論理ボリュームテーブル５１１が更新されたことを検知し、ステップＳ３３で交換されたプライマリスライスおよびセカンダリスライスそれぞれを管理する２つのストレージノードに対して更新内容を通知する。
【０１１１】
［ステップＳ３５］種別変更部５５０は、ステップＳ３２で条件を具備する全てのセグメントを選択したか否か判断する。全てのセグメントを選択した場合には、処理が終了する。未選択のセグメントが存在する場合には、処理がステップＳ３２に進められる。
【０１１２】
このようにして、種別変更部５５０は、再編成中のストレージノードに割り当てられたプライマリスライスを全てセカンダリスライスに変更する。これにより、再編成を行うことによる応答性能の低下を避けることができる。また、再編成が終了すると、その後は負荷状態に応じてプライマリスライスとセカンダリスライスとの交換が行われ、再編成が終了したストレージノードに再びアクセスの一部が振り向けられる。
【０１１３】
次に、アクセスノード６００がストレージノード１００，２００，３００，４００にアクセスする際の通信の流れの例について説明する。
図１５は、アクセスノードからのデータアクセスの流れを示すシーケンス図である。以下、図１５に示す処理をステップ番号に沿って説明する。
【０１１４】
［ステップＳ４１］アクセスノード６００は、コントロールノード５００に参照要求を送信する。
［ステップＳ４２］コントロールノード５００は、参照要求に応答して、最新の論理ボリュームをアクセスノード６００に送信する。

【０１１５】
［ステップＳ４３］アクセスノード６００は、ステップＳ４２で取得した論理ボリュームに基づいてストレージノードにアクセスする。ここでは、アクセスノード６００は、ストレージノード１００にＲｅａｄ要求を送信する。
【０１１６】
［ステップＳ４４］ストレージノード１００は、Ｒｅａｄ要求に応答して、ストレージ装置１１０からデータを取得し、取得したデータをアクセスノード６００に送信する。
［ステップＳ４５］コントロールノード５００は、ストレージノードの負荷を分散させるために、一部のセグメントのプライマリスライスとセカンダリスライスとを交換する。ここでは、コントロールノード５００は、ステップＳ４３でアクセスノード６００が読み込みを行ったデータが属するセグメントのプライマリスライスとセカンダリスライスとを交換するものとする。
【０１１７】
［ステップＳ４６ａ］コントロールノード５００は、交換前にプライマリスライスが割り当てられていたストレージノード１００に対して、セカンダリスライスに種別が変更された旨を通知する。
【０１１８】
［ステップＳ４６ｂ］コントロールノード５００は、交換前にセカンダリスライスが割り当てられていたストレージノードに対して、プライマリスライスに種別が変更された旨を通知する。ここでは、交換前にセカンダリスライスが割り当てられていたストレージノードは、ストレージノード４００であるとする。
【０１１９】
［ステップＳ４７］アクセスノード６００は、ステップＳ４２で取得した論理ボリュームに基づいて、ステップＳ４３で読み込みを行ったデータを再度読み込む。すなわち、アクセスノード６００は、ストレージノード１００にＲｅａｄ要求を送信する。
【０１２０】
［ステップＳ４８］ストレージノード１００は、Ｒｅａｄ要求で示されるデータが属するスライスがセカンダリスライスであるため、要求を拒否する旨をアクセスノード６００に応答する。
【０１２１】
［ステップＳ４９］アクセスノード６００は、コントロールノード５００に参照要求を送信する。
［ステップＳ５０］コントロールノード５００は、参照要求に応答して、最新の論理ボリュームをアクセスノード６００に送信する。
【０１２２】
［ステップＳ５１］アクセスノード６００は、ステップＳ５０で取得した論理ボリュームに基づいて、ストレージノード４００にＲｅａｄ要求を送信する。
［ステップＳ５２］ストレージノード４００は、Ｒｅａｄ要求に応答して、ストレージ装置４１０からデータを取得し、取得したデータをアクセスノード６００に送信する。
【０１２３】
このようにして、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先のストレージノードを特定する。ここで、コントロールノード５００がプライマリスライスとセカンダリスライスとの交換を行うと、アクセスノード６００はストレージノードへのアクセスに失敗する場合がある。この場合、アクセスノード６００は、コントロールノード５００から最新の論理ボリュームを取得し直し、ストレージノードへのアクセスを再度試みる。
【０１２４】
このような分散ストレージシステムによれば、アクセス用のプライマリスライスが割り当てられたストレージノードとバックアップ用のセカンダリスライスが割り当てられたストレージノードとの間で負荷が許容値を超えると、プライマリスライスとセカンダリスラ

イスとが自動的に交換される。

【０１２５】

これにより、データを移動することなく、当該データのアクセス先のストレージノードが変更される。このとき、アクセス単位でアクセス先が判定されるわけではない。従って、負荷分散の処理自体が原因で応答が遅くなることを防止し、分散ストレージシステムの応答性能を確実に高めることができる。また、データの配置管理は、ファイル単位で行われるのではなく、記憶領域の大きさに応じて分割したスライス単位で行われるため、配置管理の機構が簡潔になる。

【０１２６】

また、再編成中のストレージ装置に対しては、プライマリスライスを割り当てないようにすることで、アクセスができる限り発生しないようにすることができる。これにより、応答性能の低下を防止できる。

【０１２７】

以上、本発明のストレージ管理プログラム、ストレージ管理装置およびストレージ管理方法を図示の実施の形態に基づいて説明したが、本発明はこれに限定されるものではなく、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工程が付加されていてもよい。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。

【０１２８】

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、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。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としては、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。

【０１２９】

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の記憶装置に格納しておき、ネットワークを介して、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。

【０１３０】

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も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毎に、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。

【０１３１】

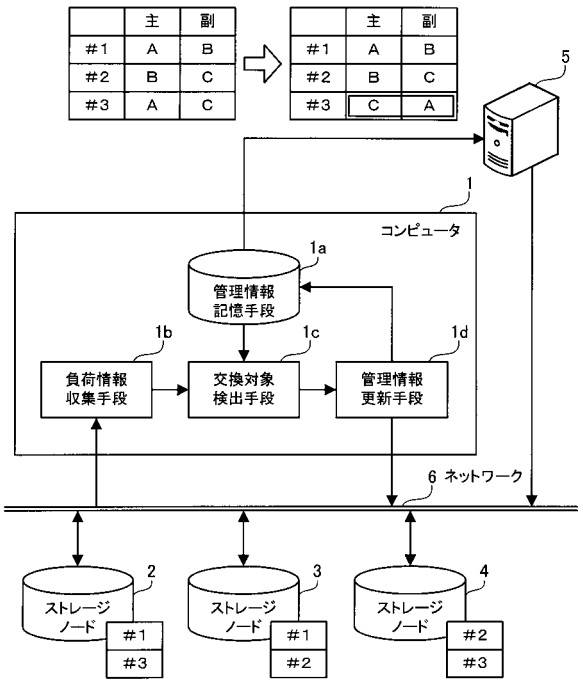
上記については単に本発明の原理を示すものである。さらに、多数の変形、変更が当業者にとって可能であり、本発明は上記に示し、説明した正確な構成および応用例に限定されるものではなく、対応するすべての変形例および均等物は、添付の請求項およびその均等物による本発明の範囲とみなされる。

【符号の説明】

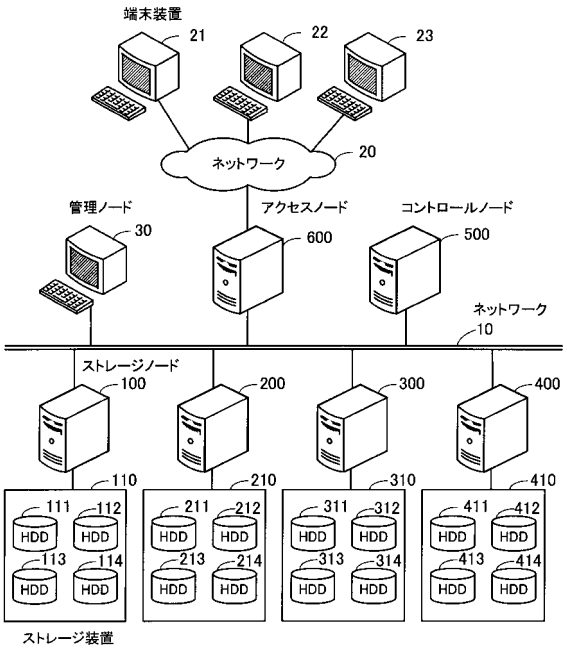
【０１３２】

１，５　コンピュータ
１ａ　管理情報記憶手段
１ｂ　負荷情報収集手段
１ｃ　交換対象検出手段
１ｄ　管理情報更新手段
２，３，４　ストレージノード
６　ネットワーク

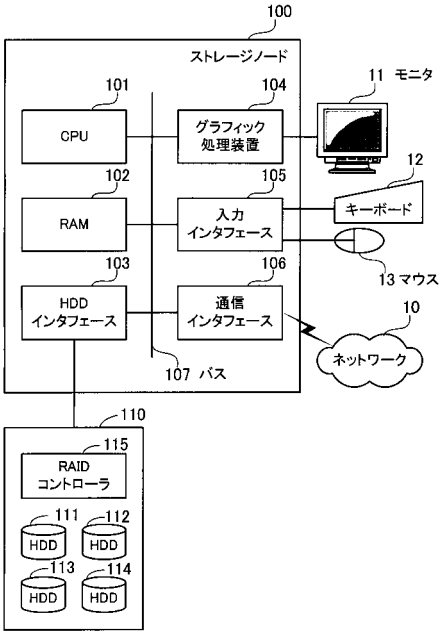
【図１】



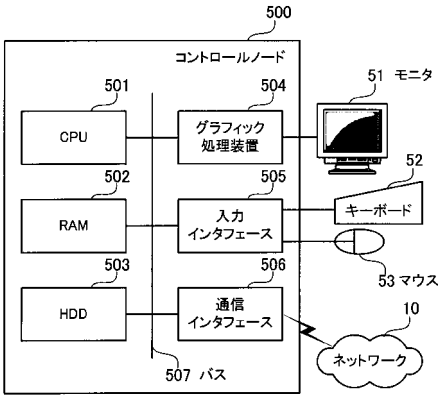
【図２】



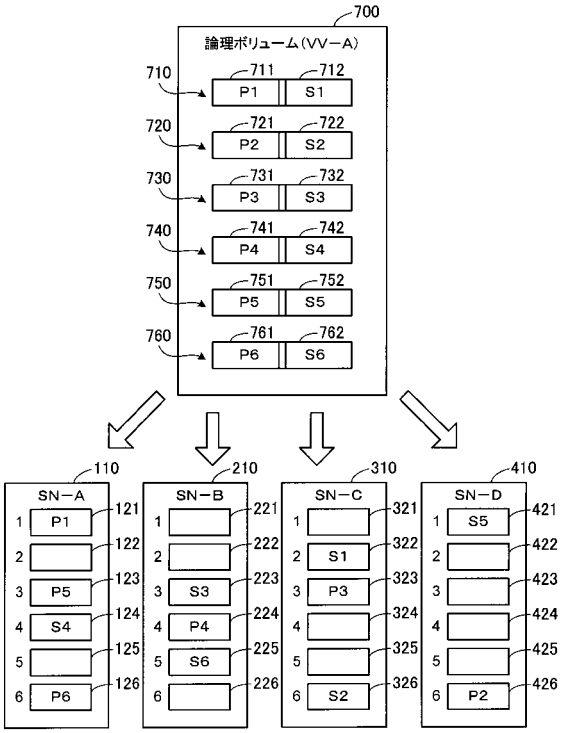
【図３】


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

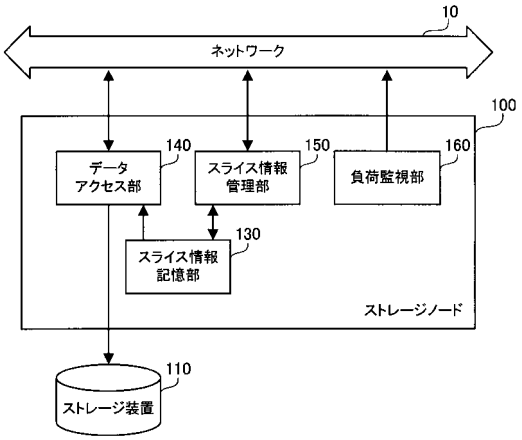
【図４】


500
コントロールノード
501
CPU
504
グラフィック処理装置
51 モニタ
502
RAM
505
入力インタフェース
52
キーボード
53 マウス
503
HDD
506
通信インタフェース
10
ネットワーク
507 パス

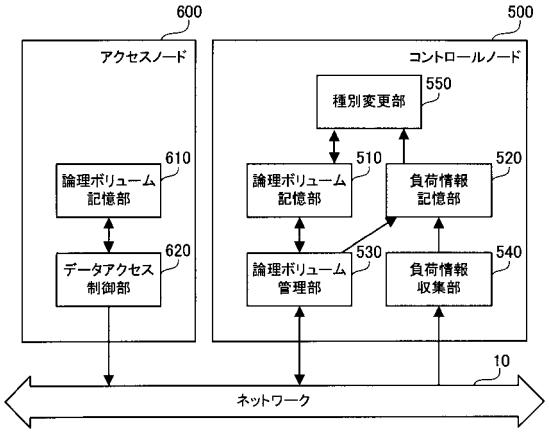
【図５】


700
論理ボリューム(VV−A)
710
711
712
P1
S1
720
721
722
P2
S2
730
731
732
P3
S3
740
741
742
P4
S4
750
751
752
P5
S5
760
761
762
P6
S6
110
SN−A
210
SN−B
310
SN−C
410
SN−D
121
122
123
124
125
126
221
222
223
224
225
226
321
322
323
324
325
326
421
422
423
424
425
426

【図６】


10
ネットワーク
100
データアクセス部
140
スライス情報管理部
150
負荷監視部
160
スライス情報記憶部
130
ストレージノード
110
ストレージ装置

【図７】



【図８】

131

スライス情報テーブル						
ディスク	物理アドレス	ブロック数	ボリューム	論理アドレス	種別	セカンダリ
sd−a	0	1024	VV−1	0	P	SN−C, sd−a, 1024
	2048	512	VV−1	4608	P	SN−D, sd−b, 0
	3072	512	VV−1	4096	S	−
	5120	1024	VV−1	5120	P	SN−B, sd−b, 4096

【図９】

511

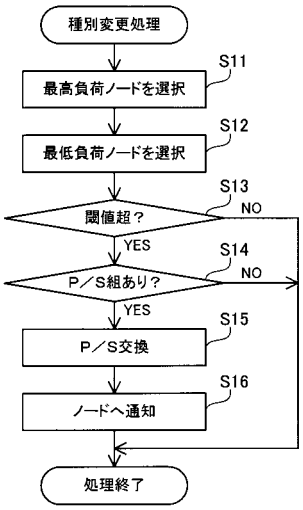
論理ボリュームテーブル（VV−1）						
セグメント	論理アドレス	ブロック数	種別	ノード	ディスク	物理アドレス
1	0	1024	P	SN−A	sd−a	0
			S	SN−C	sd−a	1024
2	1024	2048	P	SN−D	sd−b	5120
			S	SN−C	sd−a	5120
3	3072	1024	P	SN−C	sd−a	2048
			S	SN−B	sd−b	2048
4	4096	512	P	SN−B	sd−b	3072
			S	SN−A	sd−a	3072
5	4608	512	P	SN−A	sd−a	2048
			S	SN−D	sd−b	0
6	5120	1024	P	SN−A	sd−a	5120
			S	SN−B	sd−b	4096

【図１０】

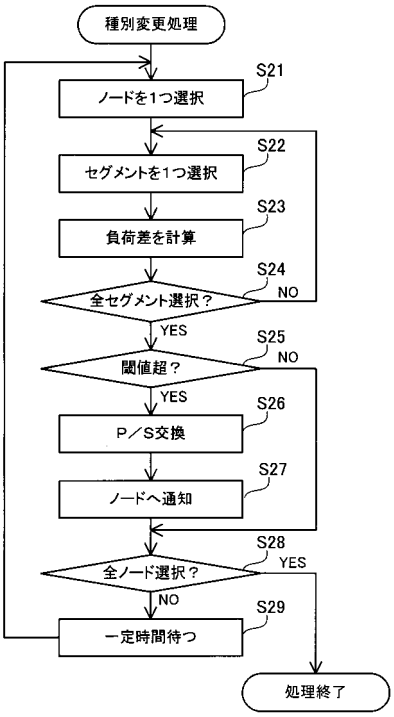
521

負荷情報テーブル				
ノード	CPU負荷	I／O数	期間I／O数	I／F使用率
SN−A	75%	10000	1000	66%
SN−B	50%	5600	897	35%
SN−C	45%	6500	1210	10%
SN−D	30%	7880	3084	40%

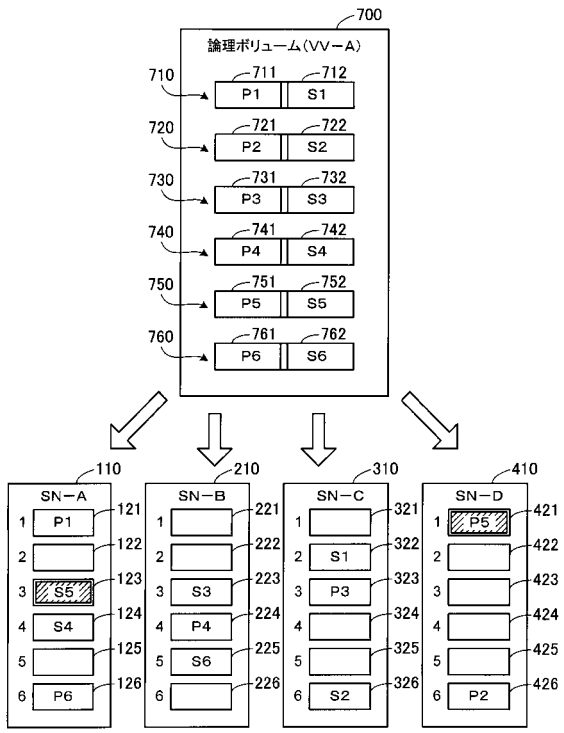
【図１１】



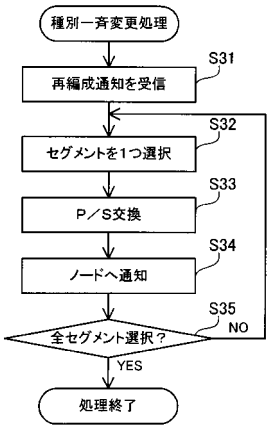
【図１２】



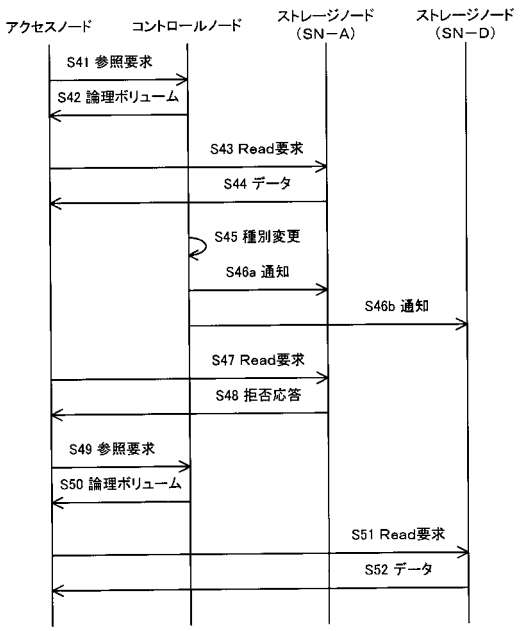
【図１３】



【図１４】



【図１５】


アクセスノード
コントロールノード
ストレージノード
（SN－A）
ストレージノード
（SN－D）
S41 参照要求
S42 論理ボリューム
S43 Read要求
S44 データ
S45 種別変更
S46a 通知
S46b 通知
S47 Read要求
S48 拒否応答
S49 参照要求
S50 論理ボリューム
S51 Read要求
S52 データ

フロントページの続き

(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　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　大江　和一
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　渡辺　高志
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　熊野　達夫
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　北村　学

(56)参考文献　特開２００６－０１２００５（ＪＰ，Ａ）
特開２００６－２２８１８８（ＪＰ，Ａ）
特開２００４－１９９２６４（ＪＰ，Ａ）
特開平０６－３３２７８２（ＪＰ，Ａ）
特開平１１－１６１５５５（ＪＰ，Ａ）
特開２００１－０５１９６３（ＪＰ，Ａ）
特開２００６－１８５４１３（ＪＰ，Ａ）
特開２００３－２９６１６７（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
G06F　3/06
G06F　12/00
JSTPlus(JDreamII)