

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4772569号
(P4772569)

(45) 発行日 平成23年9月14日 (2011.9.14)

(24) 登録日 平成23年7月1日 (2011.7.1)

(51) Int.Cl.

F I

G O 6 F 12/00 (2006.01)

G O 6 F 13/10 (2006.01)

G O 6 F 3/06 (2006.01)

G O 6 F 12/00 5 4 5 B

G O 6 F 12/00 5 4 5 A

G O 6 F 12/00 5 1 4 E

G O 6 F 13/10 3 4 O A

G O 6 F 3/06 3 O 1 Z

請求項の数 12 (全 36 頁)

(21) 出願番号 特願2006-106255 (P2006-106255)
(22) 出願日 平成18年4月7日 (2006.4.7)
(65) 公開番号 特開2007-280099 (P2007-280099A)
(43) 公開日 平成19年10月25日 (2007.10.25)
審査請求日 平成20年8月19日 (2008.8.19)

(73) 特許権者 000005108
株式会社日立製作所
東京都千代田区丸の内一丁目6番6号
(74) 代理人 100075513
弁理士 後藤 政喜
(74) 代理人 100114236
弁理士 藤井 正弘
(74) 代理人 100120260
弁理士 飯田 雅昭
(72) 発明者 揚妻 匡邦
神奈川県川崎市麻生区王禅寺1099番地
株式会社日立製作所 システム開発研究
所内

最終頁に続く

(54) 【発明の名称】 共通名前空間においてディレクトリ単位のマイグレーションを実行するシステム及び方法

(57) 【特許請求の範囲】

【請求項 1】

複数のサーバと、前記複数のサーバに接続されるディスクサブシステムと、を備えるストレージシステムにおいて、

前記複数のサーバは、第1サーバ及び第2サーバを含み、

前記各サーバは、ネットワークに接続されるインターフェースと、前記インターフェースに接続されるプロセッサと、前記プロセッサに接続されるメモリと、を備え、

前記ディスクサブシステムは、一つ以上のファイルシステムを含む一つ以上の論理デバイスを備え、

前記ファイルシステムに含まれるディレクトリのうち最上位のディレクトリには、前記ストレージシステム内で一意の第1のファイルシステム識別子が付与され、

前記ファイルシステムに含まれるディレクトリのうち前記最上位のディレクトリ以外の一つには、前記ストレージシステム内で一意の第2のファイルシステム識別子が付与され、

前記ファイルシステムに含まれるファイルには、前記ファイルシステム内で一意のファイル識別子が付与され、

前記第1サーバが備える前記プロセッサは、

前記第2のファイルシステム識別子が付与された前記ディレクトリより下位のファイルの名前解決を実行して、前記第2のファイルシステム識別子及び前記ファイル識別子を含むファイルハンドルを前記ネットワークに接続されたクライアント計算機に送信し、

10

20

前記クライアント計算機から前記第 2 のファイルシステム識別子及び前記ファイル識別子を含むファイルハンドルを指定したアクセス要求を受信すると、前記第 2 のファイルシステム識別子が付与された前記ディレクトリより下位の、前記ファイルハンドルによって指定されたファイル識別子が付与されたファイルにアクセスし、

前記第 2 のファイルシステム識別子が付与された前記ディレクトリを前記第 2 サーバへマイグレーションする指示を、前記インターフェースを介して受信すると、前記マイグレーションの対象の前記ディレクトリより下位のの前記ファイルと、前記マイグレーションの対象の前記ディレクトリに付与された前記第 2 のファイルシステム識別子と、を前記第 2 サーバへ送信し、

前記第 2 サーバが備える前記プロセッサは、前記クライアント計算機から、前記第 2 のファイルシステム識別子及び前記ファイル識別子を含むファイルハンドルを指定したアクセス要求を受信すると、前記第 2 のファイルシステム識別子が付与された前記ディレクトリより下位の、前記ファイルハンドルによって指定されたファイル識別子が付与されたファイルにアクセスすることを特徴とするストレージシステム。

【請求項 2】

前記メモリは、前記ディレクトリと、前記ディレクトリに付与された前記ファイルシステム識別子と、を対応付ける情報を格納することを特徴とする請求項 1 に記載のストレージシステム。

【請求項 3】

ストレージシステムと、クライアント計算機と、前記ストレージシステム及び前記クライアント計算機を接続するネットワークと、を備える計算機システムにおいて、

前記ストレージシステムは、複数のサーバと、前記複数のサーバに接続されるディスクサブシステムと、を備え、

前記複数のサーバは、第 1 サーバ及び第 2 サーバを含み、

前記各サーバは、前記ネットワークに接続されるインターフェースと、前記インターフェースに接続されるプロセッサと、前記プロセッサに接続されるメモリと、を備え、

前記ディスクサブシステムは、一つ以上のファイルシステムを含む一つ以上の論理デバイスを備え、

前記ファイルシステムに含まれるディレクトリのうち最上位のディレクトリには、前記ストレージシステム内で一意の第 1 のファイルシステム識別子が付与され、

前記ファイルシステムに含まれるディレクトリのうち前記最上位のディレクトリ以外の一つには、前記ストレージシステム内で一意の第 2 のファイルシステム識別子が付与され、

前記ファイルシステムに含まれるファイルには、前記ファイルシステム内で一意のファイル識別子が付与され、

前記第 1 サーバが備える前記プロセッサは、

前記第 2 のファイルシステム識別子が付与された前記ディレクトリより下位のファイルの名前解決を実行して、前記第 2 のファイルシステム識別子及び前記ファイル識別子を含むファイルハンドルを前記クライアント計算機に送信し、

前記クライアント計算機から前記第 2 のファイルシステム識別子及び前記ファイル識別子を含むファイルハンドルを指定したアクセス要求を受信すると、前記第 2 のファイルシステム識別子が付与された前記ディレクトリより下位の、前記ファイルハンドルによって指定されたファイル識別子が付与されたファイルにアクセスし、

前記第 2 のファイルシステム識別子が付与された前記ディレクトリを前記第 2 サーバへマイグレーションする指示を、前記インターフェースを介して受信すると、前記マイグレーションの対象の前記ディレクトリより下位のの前記ファイルと、前記マイグレーションの対象の前記ディレクトリに付与された前記第 2 のファイルシステム識別子と、を前記第 2 サーバへ送信し、

前記第 2 サーバが備える前記プロセッサは、前記クライアント計算機から、前記第 2 のファイルシステム識別子及び前記ファイル識別子を含むファイルハンドルを指定したアク

10

20

30

40

50

セス要求を受信すると、前記第2のファイルシステム識別子が付与された前記ディレクトリより下位の、前記ファイルハンドルによって指定されたファイル識別子が付与されたファイルにアクセスすることを特徴とする計算機システム。

【請求項4】

前記メモリは、前記ディレクトリと、前記ディレクトリに付与された前記ファイルシステム識別子と、を対応付ける情報を格納することを特徴とする請求項3に記載の計算機システム。

【請求項5】

前記計算機システムは、さらに管理計算機を備え、

前記管理計算機は、前記ネットワークに接続されるインターフェースと、前記インターフェースに接続されるプロセッサと、前記プロセッサに接続されるメモリと、を備え、

前記計算機システムは、前記ディレクトリに既に付与された前記ファイルシステム識別子の値を管理する管理情報を保持し、

前記管理計算機が備える前記プロセッサは、前記管理情報を参照して、まだいずれの前記ディレクトリにも付与されていない値を、まだ前記ファイルシステム識別子を付与されていない前記ディレクトリに、前記ファイルシステム識別子として付与することを特徴とする請求項3に記載の計算機システム。

【請求項6】

前記計算機システムは、さらに管理計算機を備え、

前記管理計算機は、前記ネットワークに接続されるインターフェースと、前記インターフェースに接続されるプロセッサと、前記プロセッサに接続されるメモリと、を備え、

前記管理計算機が備える前記プロセッサは、ファイルアクセスに関する統計情報に基づいて、前記ディレクトリを対象とするマイグレーションを実行するか否かを判定することを特徴とする請求項3に記載の計算機システム。

【請求項7】

前記統計情報は、前記各サーバが備えるプロセッサの使用率であり、

前記管理計算機が備える前記プロセッサは、前記第1サーバが備えるプロセッサの使用率が所定の閾値を越えた場合、前記第1サーバが管理する前記ファイルシステムに含まれ、かつ、前記ファイルシステム識別子が付与された前記ディレクトリを選択し、前記選択されたディレクトリを対象とするマイグレーションを実行する指示を前記第1サーバに送信することを特徴とする請求項6に記載の計算機システム。

【請求項8】

ストレージシステムと、クライアント計算機と、前記ストレージシステム及び前記クライアント計算機を接続するネットワークと、を備える計算機システムの制御方法であって、

前記ストレージシステムは、第1サーバ及び第2サーバを含む複数のサーバと、前記複数のサーバに接続されるディスクサブシステムと、を備え、

前記各サーバは、前記ネットワークに接続されるインターフェースと、前記インターフェースに接続されるプロセッサと、前記プロセッサに接続されるメモリと、を備え、

前記ディスクサブシステムは、一つ以上のファイルシステムを含む一つ以上の論理デバイスを備え、

前記ファイルシステムに含まれるディレクトリのうち最上位のディレクトリには、前記ストレージシステム内で一意の第1のファイルシステム識別子が付与され、

前記ファイルシステムに含まれるディレクトリのうち前記最上位のディレクトリ以外の一つには、前記ストレージシステム内で一意の第2のファイルシステム識別子が付与され、

前記ファイルシステムに含まれるファイルには、前記ファイルシステム内で一意のファイル識別子が付与され、

前記方法は、

前記第1サーバが備える前記プロセッサが、前記第2のファイルシステム識別子が付与

10

20

30

40

50

された前記ディレクトリより下位のファイルの名前解決を実行して、前記第2のファイルシステム識別子及び前記ファイル識別子を含むファイルハンドルを前記クライアント計算機に送信し、

前記第1サーバが備える前記プロセッサが、前記クライアント計算機から前記第2のファイルシステム識別子及び前記ファイル識別子を含むファイルハンドルを指定したアクセス要求を受信すると、前記第2のファイルシステム識別子が付与された前記ディレクトリより下位の、前記ファイルハンドルによって指定されたファイル識別子が付与されたファイルにアクセスし、

前記第1サーバが備える前記プロセッサが、前記第2のファイルシステム識別子が付与された前記ディレクトリを前記第2サーバへマイグレーションする指示を、前記インターフェースを介して受信すると、前記マイグレーションの対象の前記ディレクトリより下位の前記ファイルと、前記マイグレーションの対象の前記ディレクトリに付与された前記第2のファイルシステム識別子と、を前記第2サーバへ送信し、

前記第2サーバが備える前記プロセッサが、前記クライアント計算機から、前記第2のファイルシステム識別子及び前記ファイル識別子を含むファイルハンドルを指定したアクセス要求を受信すると、前記第2のファイルシステム識別子が付与された前記ディレクトリより下位の、前記ファイルハンドルによって指定されたファイル識別子が付与されたファイルにアクセスすることを特徴とする方法。

【請求項9】

前記メモリは、前記ディレクトリと、前記ディレクトリに付与された前記ファイルシステム識別子と、を対応付ける情報を格納することを特徴とする請求項8に記載の方法。

【請求項10】

前記計算機システムは、さらに管理計算機を備え、

前記管理計算機は、前記ネットワークに接続されるインターフェースと、前記インターフェースに接続されるプロセッサと、前記プロセッサに接続されるメモリと、を備え、

前記計算機システムは、前記ディレクトリに既に付与された前記ファイルシステム識別子の値を管理する管理情報を保持し、

前記方法は、前記管理計算機が備える前記プロセッサが、前記管理情報を参照して、まだいずれの前記ディレクトリにも付与されていない値を、まだ前記ファイルシステム識別子を付与されていない前記ディレクトリに、前記ファイルシステム識別子として付与することを特徴とする請求項8に記載の方法。

【請求項11】

前記計算機システムは、さらに管理計算機を備え、

前記管理計算機は、前記ネットワークに接続されるインターフェースと、前記インターフェースに接続されるプロセッサと、前記プロセッサに接続されるメモリと、を備え、

前記方法は、前記管理計算機が備える前記プロセッサが、ファイルアクセスに関する統計情報に基づいて、前記ディレクトリを対象とするマイグレーションを実行するか否かを判定することを特徴とする請求項8に記載の方法。

【請求項12】

前記統計情報は、前記各サーバが備えるプロセッサの使用率であり、

前記方法は、前記管理計算機が備える前記プロセッサが、前記第1サーバが備えるプロセッサの使用率が所定の閾値を越えた場合、前記第1サーバが管理する前記ファイルシステムに含まれ、かつ、前記ファイルシステム識別子が付与された前記ディレクトリを選択し、前記選択されたディレクトリを対象とするマイグレーションを実行する指示を前記第1サーバに送信することを特徴とする請求項11に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ネットワークを介してクライアントと接続され、クライアントが用いるデータを記憶する、ファイルサーバ又はネットワークアタッチドストレージ (Network Attach

10

20

30

40

50

ed Storage、N A S) において、複数のサーバに負荷を分散するための技術に関する。

【背景技術】

【 0 0 0 2 】

ネットワークに接続されたストレージシステムが、そのネットワークに接続されたクライアント計算機の共有ディスクとして使用される、ネットワークアタッチドストレージ (Network Attached Storage、N A S) が提案されている。N A S は、ネットワークインターフェース等を含むサーバと、データを格納するディスクドライブとによって構成される。

【 0 0 0 3 】

例えば、特許文献 1 には、複数のサーバを備え、各サーバがネットワークに接続されるクラスタ構成の N A S が開示されている。特許文献 1 に記載されたシステムにおいて、ネットワークエレメント、スイッチングエレメント及びディスクエレメントが、N A S のサーバに相当する。特許文献 1 に記載されたシステムは、ファイルシステムを共有する複数のネットワークエレメントを備えることができる。また、特許文献 1 に記載されたシステムは、ディスクエレメントを複数備えることによって、ディスク単位でデータをマイグレーションすることができる。さらに、各ネットワークエレメントがディスクエレメントによって管理される全てのファイルシステムにアクセスすることができる。ディスクエレメント間で、ファイルシステムを格納しているディスクのマイグレーションが実行された場合も、各ネットワークエレメントは、マイグレーションによる影響を受けることなく、全てのファイルシステムにアクセスをすることができる。

【 0 0 0 4 】

ネットワーク上に分散して存在するファイルにアクセスするためのファイルシステムの一つとして、ネットワークファイルシステム (N F S) が提案されている。N F S の現時点の最新版である N F S v 4 (RFC3530) (非特許文献 1 参照) によれば、ファイルシステムがサーバ間でマイグレーションされた場合、サーバは、クライアントからの当該ファイルシステムへのアクセスに対して、マイグレーション先の位置の情報を通知する。この通知を受けたクライアントは、通知された位置情報に従って、マイグレーション先のファイルシステムにアクセスすることができる。

【特許文献 1】米国特許第 6 6 7 1 7 7 3 号明細書

【非特許文献 1】RFC3530、NFS version4、P. 57 - 61、IETF Home Page、[online]、[平成 1 8 年 3 月 1 5 日検索]、インターネット < URL: <http://www.ietf.org/home.html> >

【発明の開示】

【発明が解決しようとする課題】

【 0 0 0 5 】

複数のサーバが管理する複数のファイルシステムを、一つの名前空間によってクライアントに提供することができる。そのような名前空間は、グローバルネームスペース (G N S) と呼ばれる。G N S を適用したシステムでは、負荷の高いサーバから負荷の低いサーバへファイルを転送するマイグレーションによって、複数のサーバ間の負荷の均衡を図ることができる。

【 0 0 0 6 】

G N S において、負荷均衡のためのマイグレーションをクライアントから透過的に実行するために、ファイルに付与された I D (ファイルハンドル) を維持する必要がある。言い換えると、マイグレーションの前に付与されたファイルハンドルが、マイグレーションの後にも使用できることが必要である。なお、ファイルハンドルは、各サーバが管理するファイルシステムをそのサーバ内で一意に識別するファイルシステム I D と、各ファイルをファイルシステム内で一意に識別するファイル I D とを含む。

【 0 0 0 7 】

ファイルハンドルを維持するために、システム内で一意に定まるファイルシステム I D をファイルシステムに割り振り、ファイルシステム単位でマイグレーションを実行しても

10

20

30

40

50

よい。この場合、ファイルシステムを管理しているサーバに依存しない、システム内で一意のファイルハンドルが付与される。その結果、クライアントは、自分がアクセスしようとするファイルが、どのサーバによって管理されているファイルシステムに属するかを意識する必要がない。すなわち、ファイルシステムが一つのサーバから別のサーバにマイグレーションされた後も、クライアントは、マイグレーションが実行される前に付与されたファイルハンドルを使用して、そのファイルシステムに属するファイルにアクセスすることができる。このようにして、クライアントから透過的なファイルシステムのマイグレーションが実現される。

【 0 0 0 8 】

しかし、サーバ間の負荷の均衡を目的としてマイグレーションを行う場合、ファイルシステム単位では粒度が粗いため、十分な負荷均衡が行えない場合がある。例えば、一つのファイルシステムの中に負荷が集中している場合、そのファイルシステム全体をマイグレーションしても、そのファイルシステムの負荷全体が移動するだけである。その結果、負荷を複数のサーバに分散して均衡させることができない。この場合、ファイルシステムより細かい単位で、例えば、ディレクトリ単位でのマイグレーションを実行する必要がある。

【 0 0 0 9 】

例えば、あるサーバが管理している任意のファイル又はディレクトリを、別のサーバが管理するファイルシステムにコピーすることによって、ファイルシステムより細かい単位のマイグレーションを実現することができる。その結果、そのファイル等に対するアクセスによる負荷を分散させることによって、複数のサーバの負荷を均衡させることができる。しかし、この場合、クライアントは、マイグレーションの実行前に取得したファイルハンドルを用いて、マイグレーションされたファイルにアクセスすることができない。すなわち、再度名前解決を実行する必要がある。

【 0 0 1 0 】

ファイルシステムよりも細かい単位で、クライアントから透過的なマイグレーションを実行する方法については、開示されていない。

【課題を解決するための手段】

【 0 0 1 1 】

本願で開示する代表的な発明は、複数のサーバと、前記複数のサーバに接続されるディスクサブシステムと、を備えるストレージシステムにおいて、前記複数のサーバは、第1サーバ及び第2サーバを含み、前記各サーバは、ネットワークに接続されるインターフェースと、前記インターフェースに接続されるプロセッサと、前記プロセッサに接続されるメモリと、を備え、前記ディスクサブシステムは、一つ以上のファイルシステムを含む一つ以上の論理デバイスを備え、前記ファイルシステムに含まれるディレクトリのうち最上位のディレクトリには、前記ストレージシステム内で一意の第1のファイルシステム識別子が付与され、前記ファイルシステムに含まれるディレクトリのうち前記最上位のディレクトリ以外の一つには、前記ストレージシステム内で一意の第2のファイルシステム識別子が付与され、前記ファイルシステムに含まれるファイルには、前記ファイルシステム内で一意のファイル識別子が付与され、前記第1サーバが備える前記プロセッサは、前記第2のファイルシステム識別子が付与された前記ディレクトリより下位のファイルの名前解決を実行して、前記第2のファイルシステム識別子及び前記ファイル識別子を含むファイルハンドルを前記ネットワークに接続されたクライアント計算機に送信し、前記クライアント計算機から前記第2のファイルシステム識別子及び前記ファイル識別子を含むファイルハンドルを指定したアクセス要求を受信すると、前記第2のファイルシステム識別子が付与された前記ディレクトリより下位の、前記ファイルハンドルによって指定されたファイル識別子が付与されたファイルにアクセスし、前記第2のファイルシステム識別子が付与された前記ディレクトリを前記第2サーバへマイグレーションする指示を、前記インターフェースを介して受信すると、前記マイグレーションの対象の前記ディレクトリより下位の前記ファイルと、前記マイグレーションの対象の前記ディレクトリに付与された前記

第 2 のファイルシステム識別子と、を前記第 2 サーバへ送信し、前記第 2 サーバが備える前記プロセッサは、前記クライアント計算機から、前記第 2 のファイルシステム識別子及び前記ファイル識別子を含むファイルハンドルを指定したアクセス要求を受信すると、前記第 2 のファイルシステム識別子が付与された前記ディレクトリより下位の、前記ファイルハンドルによって指定されたファイル識別子が付与されたファイルにアクセスすることを特徴とする。

【 0 0 1 2 】

なお、望ましくは、前記プロセッサは、前記第 2 識別子が付与された前記ディレクトリより下位のファイルを、当該ファイルより上位の前記第 2 識別子が付与されたディレクトリのうち当該ファイルに最も近いディレクトリに付与された前記第 2 識別子と、前記第 3 識別子と、によって識別する。

10

【発明の効果】

【 0 0 1 3 】

本発明の一形態によれば、ファイルシステムを分割し、分割されたファイルシステムの部分ごとにマイグレーションを実行することができる。このため、クライアントから透過的に、ディレクトリ単位のマイグレーションを実行することができる。このため、きめ細かい単位でマイグレーションを実行することができ、サーバ間のより良い負荷均衡を実現できる。その結果、システム全体の性能向上が実現される。

【発明を実施するための最良の形態】

【 0 0 1 4 】

20

図 1 は、本発明の実施の形態のストレージシステム 1 0 0 の構成を示すブロック図である。

【 0 0 1 5 】

本実施の形態の計算機システムは、ストレージシステム 1 0 0、管理計算機 1 4 0、クライアント 1 5 0 A、クライアント 1 5 0 B 及びこれらを相互に接続する L A N 1 6 0 を備える。

【 0 0 1 6 】

管理計算機 1 4 0 は、ストレージシステム 1 0 0 を管理するための処理を実行する計算機である。例えば、管理計算機 1 4 0 は、ストレージシステム 1 0 0 内のファイルシステムの作成、ファイルシステムのマウント又はマイグレーションをサーバ 1 1 0 に指示する。さらに、管理計算機 1 4 0 は、ディスクサブシステム 1 2 0 及びスイッチ 1 3 0 に設定変更を指示する。

30

【 0 0 1 7 】

管理計算機 1 4 0 は、少なくとも、相互に接続された入力装置 1 4 1、管理画面 1 4 2、ディスク 1 4 3、C P U 1 4 4、ローカルメモリ 1 4 5 及びネットワークインターフェース (I / F) 1 4 6 を備える。管理計算機 1 4 0 は、C P U 1 4 4、ローカルメモリ 1 4 5 及びネットワーク I / F 1 4 6 を備える筐体 (図示省略) と、その筐体に接続された入力装置 1 4 1、管理画面 1 4 2 及びディスク 1 4 3 によって構成されてもよい。入力装置 1 4 1 は、例えば、システム管理者によって使用されるキーボード又はポインティングデバイスである。管理画面 1 4 2 は、例えば、システム管理者に情報を表示する画像表示装置である。管理画面 1 4 2 に表示される内容及びポインティングデバイスによる操作については、後述する (図 1 6 及び図 1 7 参照) 。

40

【 0 0 1 8 】

ディスク 1 4 3 には、サーバ 1 1 0 等と通信するためのプログラム、及び、ディスクサブシステムを管理するためのプログラム等が格納される。これらのプログラムは、ローカルメモリ 1 4 5 上に読み出され、C P U 1 4 4 によって実行される。ネットワーク I / F 1 4 6 は、サーバ 1 1 0 及びディスクサブシステム 1 2 0 との通信に使用される。

【 0 0 1 9 】

クライアント 1 5 0 A 及び 1 5 0 B は、ストレージシステム 1 0 0 のファイルにアクセスする計算機である。クライアント 1 5 0 A 及び 1 5 0 B は、ストレージシステム 1 0 0

50

にファイルを書き込み、又は、ストレージシステム 100 からファイルを読み出す。このとき、ストレージシステム 100 が管理するファイルシステムが使用される。クライアント 150 A 及び 150 B は、少なくとも、ファイルアクセスを実行するプログラムを格納するメモリ（図示省略）、及び、メモリに格納されたプログラムを実行するプロセッサ（図示省略）を備える。

【0020】

図 1 では、二つのクライアント（150 A 及び 150 B）を示すが、任意の数のクライアントが LAN 160 に接続され、ストレージシステム 100 にアクセスすることができる。以下の説明において、二つのクライアントを特に区別する必要がない場合、これらを総称してクライアント 150 と記載する。

10

【0021】

LAN 160 は、TCP/IP 等のプロトコルで通信するネットワークである。

【0022】

ストレージシステム 100 は、ネットワークアタッチドストレージ（Network Attached Storage、NAS）である。ストレージシステム 100 は、サーバ 110 A、サーバ 110 B、ディスクサブシステム 120 及びこれらを相互に接続するスイッチ 130 を備える。

【0023】

スイッチ 130 は、例えばファイバーチャネル（FC）のスイッチ 130 である。ストレージシステム 100 は、複数のスイッチ 130 を備えてもよい。一つ以上のスイッチ 130 がストレージエリアネットワーク（SAN）を構成していてもよい。あるいは、スイッチ 130 は、LAN のスイッチ又はストレージシステム専用のスイッチであってもよい。

20

【0024】

サーバ 110 A 及び 110 B は、クライアント 150 から受信したアクセス要求に従ってディスクサブシステム 120 にアクセスする計算機である。

【0025】

以下の説明において、サーバ 110 A 及び 110 B を特に区別する必要がない場合、これらを総称してサーバ 110 と記載する。また、以下の説明において、サーバ 110 A 及び 110 B は、単にサーバ A 及びサーバ B と記載される。図 1 では、二つのサーバ 110 A 及び 110 B を示すが、ストレージシステム 100 は、さらに多くのサーバ 110 を備えてもよい。サーバ 110 は、NAS ヘッド又は NAS ノードとも呼ばれる。複数のサーバ 110 がクラスタ構成されていてもよい。

30

【0026】

サーバ 110 A は、相互に接続されたネットワークインターフェース 111 A、CPU 112 A、ローカルメモリ 113 A 及びアダプタ 116 A を備える。

【0027】

ネットワークインターフェース 111 A は、LAN 160 に接続され、管理計算機 140 及びクライアント 150 と通信するインターフェースである。

【0028】

CPU 112 A は、サーバ 110 A の動作を制御するプロセッサである。具体的には、CPU 112 A は、ローカルメモリ 113 A に格納されたプログラムを実行する。

40

【0029】

ローカルメモリ 113 A は、例えば、半導体メモリであり、CPU 112 A が実行するプログラム及び CPU 112 A によって参照されるデータを格納する。具体的には、ローカルメモリ 113 A は、マウントテーブル（図示省略）、後述するサーバソフトウェア 200 及びマイグレーションポイント管理表 600 等を格納する。

【0030】

アダプタ 116 A は、スイッチ 130 に接続され、ディスクサブシステム 120 と通信するインターフェースである。

50

【 0 0 3 1 】

サーバ 1 1 0 B は、サーバ 1 1 0 A と同様、ネットワークインターフェース 1 1 1 B、CPU 1 1 2 B、ローカルメモリ 1 1 3 B 及びアダプタ 1 1 6 B を備える。これらは、それぞれ、ネットワークインターフェース 1 1 1 A、CPU 1 1 2 A、ローカルメモリ 1 1 3 A 及びアダプタ 1 1 6 A と同様であるため、説明を省略する。さらに多くのサーバ 1 1 0 が設けられる場合、各サーバ 1 1 0 の構成は、サーバ 1 1 0 A と同様であってよい。

【 0 0 3 2 】

ストレージシステム 1 0 0 が複数のサーバ 1 1 0 を備える場合、これらのサーバ 1 1 0 は、サーバ間通信路 1 3 5 によって相互に接続される。サーバ 1 1 0 は、サーバ間通信路 1 3 5 を経由して相互に通信することができる。

10

【 0 0 3 3 】

具体的には、G N S 管理表 5 0 0 が更新された場合に、その更新内容がサーバ間通信路 1 3 5 を経由して他のサーバ 1 1 0 に送信される。送信先のサーバ 1 1 0 は、受信した更新の内容をそのサーバ 1 1 0 内の G N S 管理表 5 0 0 に反映する。

【 0 0 3 4 】

本実施の形態のストレージシステム 1 0 0 は、図 1 に示すように、スイッチ 1 3 0 及び LAN 1 6 0 のどちらからも独立したサーバ間通信路 1 3 5 を備える。しかし、サーバ 1 1 0 は、スイッチ 1 3 0 又は LAN 1 6 0 を経由して相互に通信してもよい。あるいは、サーバ 1 1 0 は、ディスクサブシステム 1 2 0 のディスクキャッシュ 1 2 2 を使用して相互に通信してもよい。サーバ 1 1 0 が上記のいずれの経路で通信しても、本発明を実施することができる。

20

【 0 0 3 5 】

ディスクサブシステム 1 2 0 は、相互に接続されたディスクコントローラ 1 2 1 A、ディスクコントローラ 1 2 1 B、ディスクキャッシュ 1 2 2 及びディスクドライブ 1 2 3 を備える。ディスクコントローラ 1 2 1 A は、スイッチ 1 3 0 に接続されるポート 1 2 5 A 及び 1 2 5 B を備える。ディスクコントローラ 1 2 1 B は、スイッチ 1 3 0 に接続されるポート 1 2 5 C 及び 1 2 5 D を備える。以下の説明において、ディスクコントローラ 1 2 1 A 及び 1 2 1 B を特に区別する必要がない場合、これらを総称してディスクコントローラ 1 2 1 と記載する。また、ポート 1 2 5 A から 1 2 5 D を特に区別する必要がない場合、これらを総称してポート 1 2 5 と記載する。ディスクサブシステム 1 2 0 は、一つ又は三つ以上のディスクコントローラ 1 2 1 を備えてもよい。各ディスクコントローラ 1 2 1 は、一つ又は三つ以上のポート 1 2 5 を備えてもよい。

30

【 0 0 3 6 】

ディスクコントローラ 1 2 1 は、ポート 1 2 5 に接続されたスイッチ 1 3 0 を介してサーバ 1 1 0 と通信し、ディスクサブシステム 1 2 0 を制御する。具体的には、ディスクコントローラ 1 2 1 は、サーバ 1 1 0 からの要求に従ってディスクドライブ 1 2 3 にデータを書き込み、又は、ディスクドライブ 1 2 3 からデータを読み出す。

【 0 0 3 7 】

ディスクキャッシュ 1 2 2 は、例えば、半導体メモリであり、ディスクドライブ 1 2 3 に書き込まれるデータ又はディスクドライブ 1 2 3 から読み出されたデータを一時的に格納する。

40

【 0 0 3 8 】

ディスクドライブ 1 2 3 A から 1 2 3 D は、データを格納するハードディスクドライブである。ディスクサブシステムは、任意の数のディスクドライブ 1 2 3 A 等を備える。以下の説明において、ディスクドライブ 1 2 3 A から 1 2 3 D を特に区別する必要がない場合、これらを総称してディスクドライブ 1 2 3 と記載する。これらのディスクドライブ 1 2 3 は、Redundant Arrays of Inexpensive Disks (R A I D) を構成してもよい。

【 0 0 3 9 】

ディスクドライブ 1 2 3 の記憶領域は、任意の数の論理デバイス (L D E V) に分割して使用される。図 1 には、例として、4 個の L D E V 1 2 4 A から 1 2 4 D を示す。以下

50

の説明において、LDEV124Aから124Dを特に区別する必要がない場合、これらを総称してLDEV124と記載する。LDEV124は、ディスクコントローラ121によって論理的なディスクドライブとして扱われる領域である。ディスクドライブ123がRAIDを構成する場合、図1に示すように、一つのLDEV124が複数のディスクドライブ123の記憶領域によって構成されてもよい。各LDEV124の大きさは任意である。

【0040】

各LDEVには、LDEV識別子(ID)が付与される。図1の例において、LDEV124A~LDEV124DのIDは、それぞれL0~L3である。

【0041】

サーバ110は、各LDEV124内にファイルシステムを作成し、そのファイルシステムを管理する。各ファイルシステムには、ストレージシステム内で一意となるファイルシステム識別子(ID)が付与される。

【0042】

ストレージシステム100は、複数のディスクサブシステム120を備えてもよい。その場合、あるサーバ110は、スイッチ130を経由して、特定のディスクサブシステム120又は特定のディスクサブシステム120の集合のみにアクセスできてもよい。あるいは、各サーバ110が全てのディスクサブシステム120にアクセスすることができてもよい。

【0043】

スイッチ130及びディスクサブシステム120は、それぞれ、LAN160に接続される管理用ポート131及び126を備える。管理計算機140は、LAN160を経由してこれらの管理用ポート131及び126と通信することによって、スイッチ130及びディスクサブシステム120の設定を参照し、又は、更新することができる。

【0044】

図2は、本発明の実施の形態のサーバソフトウェア200の説明図である。

【0045】

サーバソフトウェア200は、CPU112が実行するプログラムを含む。具体的には、サーバソフトウェア200は、ネットワーク処理プログラム201、ファイルシステム処理プログラム202、ディスクアクセス処理プログラム203、GNS管理処理プログラム204、サーバ管理処理プログラム205、サーバ間通信処理プログラム206及びマイグレーション処理プログラム207を含む。

【0046】

ネットワーク処理プログラム201は、LAN160を介した管理計算機140及びクライアント150との通信を制御するプログラムである。

【0047】

ファイルシステム処理プログラム202は、クライアント150からファイルシステム内のファイルへのアクセス要求、及び、管理計算機140からのファイルシステムに対する指示等を処理するプログラムである。ファイルシステム処理プログラム202のより詳細な説明は後述する。

【0048】

ディスクアクセスプログラム203は、ファイルシステム処理プログラム202が受けたファイルへのアクセス要求に基づき、実際にディスクサブシステム120へアクセスするプログラムである。

【0049】

GNS管理処理プログラム204は、マイグレーション処理プログラム200がファイルシステムのマイグレーションを実行したときに、GNS管理表500を更新する。GNS管理表500の詳細は後述する(図5参照)。

【0050】

サーバ管理処理プログラム205は、管理計算機140と通信し、サーバ110の設定

10

20

30

40

50

をするプログラムである。例えば、サーバ管理処理プログラム 205 は、管理計算機 140 からファイルシステムの新規作成の指示を受けると、その指示をファイルシステム処理プログラム 202 に伝え、ファイルシステムの新規作成を実行させる。管理計算機 140 からマイグレーションの指示を受けると、サーバ管理処理プログラム 205 は、その指示をマイグレーション処理プログラム 207 に伝え、マイグレーションを実行させる。

【0051】

サーバ間通信処理プログラム 206 は、サーバ間通信路 135 を経由したサーバ 110 間の通信を制御するプログラムである。例えば、サーバ間通信処理プログラム 206 は、GNS 管理表 500 の内容を更新するため、他のサーバ 110 において実行されている GNS 管理処理プログラム 204 に更新内容を送信する。

10

【0052】

マイグレーション処理プログラム 207 は、マイグレーションを実行するプログラムである。マイグレーションとは、あるファイルシステムを、現在管理しているサーバ 110 から、別の指定されたサーバ 110 の管理下に移動させる処理のことである。

【0053】

クライアント 150 がファイルシステム内のファイルにアクセスする場合、アクセス要求は、そのファイルシステムを管理するサーバ 110 によって処理される。従って、ファイルシステムをマイグレーションすることは、マイグレーション先のサーバへ負荷を分散することになる。マイグレーション処理によって、各サーバ 110 にかかる負荷が均衡し、ストレージシステム 100 全体の性能向上が可能となる。

20

【0054】

ファイルシステム処理プログラム 202 は、マイグレーションポイント設定処理モジュール 211、ファイルアクセス処理モジュール 212、ファイルシステム作成処理モジュール 213 及びファイルシステム分割処理モジュール 214 を含む。

【0055】

マイグレーションポイント設定処理モジュール 211 は、管理用計算機 140 の指示に従って、サーバ 110 が持つマイグレーションポイント管理表 600 の更新処理を実行する。

【0056】

ファイルアクセス処理モジュール 212 は、クライアント 150 からの要求に基づき、ディレクトリ名又はファイル名の名前解決処理を実行し、ファイルの識別子であるファイルハンドルをクライアントに返す。また、例えば、クライアント 150 の要求するディレクトリ名やファイル名が、他のサーバ 110 によって管理されるディレクトリツリーに属する場合、そのディレクトリツリーを管理するサーバ 110 の位置情報（例えば、サーバ 110 の ID）をクライアント 150 へ返す。ディレクトリツリーとは、ファイルシステム全体又はその一部に相当する一つ以上のディレクトリである。

30

【0057】

ファイルシステム作成処理モジュール 213 は、管理計算機 140 の指示に従って、ファイルシステムを新規作成する。また、ファイルシステム作成処理モジュール 213 は、ファイルシステム分割処理モジュール 214 の指示に従って、ファイルシステムを新規作成する。

40

【0058】

ファイルシステム分割処理モジュール 214 は、マイグレーション処理プログラム 207 の指示に従って、ファイルシステムの分割処理を実行する。ファイルシステムの分割処理の詳細は後述する（図 9 参照）。

【0059】

図 3 は、本発明の実施の形態においてクライアント 150 に提供される共通名前空間の一例を示す説明図である。

【0060】

共通名前空間は、一つ以上のサーバ 110 によって構成される。具体的には、共通名前

50

空間は、各サーバ 1 1 0 が管理しているファイルシステムを統合することによって構成される一つの共通名前空間である。この共通名前空間をグローバルネームスペース (G N S) と呼ぶ。図 3 には、実施例を説明するために必要最小限のディレクトリ及びファイルを含む G N S 3 0 0 を示す。実際の G N S は、さらに多くのディレクトリ及びファイルを含んでもよい。

【 0 0 6 1 】

G N S 3 0 0 は、二つのサーバ、すなわち、サーバ A 及びサーバ B によって G N S が構成される。次に、各サーバの名前空間について説明する。

【 0 0 6 2 】

図 4 は、本発明の実施の形態の各サーバ 1 1 0 の名前空間の説明図である。

10

【 0 0 6 3 】

図 4 は、例として、サーバ A の名前空間 4 0 0 A 及びサーバ B の名前空間 4 0 0 B を示す。最初に、名前空間 4 0 0 A 及び 4 0 0 B を構成する方法を説明し、次に、G N S 3 0 0 を構成する方法を説明する。

【 0 0 6 4 】

サーバ A は、三つのファイルシステム 4 1 1、4 0 1 及び 4 0 2 を管理する。それぞれのファイルシステムに属するディレクトリツリーを統合することによって、名前空間 4 0 0 A が構成される。

【 0 0 6 5 】

ファイルシステム 4 1 1 に相当するディレクトリツリーの最上位ディレクトリの下に、ディレクトリ「gns」4 1 4 及び「mnt」がある。ディレクトリ「gns」4 1 4 の下に、「a」4 0 3、「b」4 0 4 及び「c」4 0 5 の三つのディレクトリがある。一方、ディレクトリ「mnt」の下には、ディレクトリ「fs0」4 1 3 及びディレクトリ「fs1」4 0 8 がある。

20

【 0 0 6 6 】

ファイルシステム 4 0 1 に相当するディレクトリツリーの最上位ディレクトリの下に、ディレクトリ「aa」4 0 6 及びディレクトリ「ab」がある。ディレクトリ「aa」の下には、ファイル「file1」がある。また、ディレクトリ「ab」の下には、ディレクトリ「ac」4 0 7 がある。ディレクトリ「ac」4 0 7 の下に、ディレクトリ「ad」及びディレクトリ「ae」がある。

30

【 0 0 6 7 】

ファイルシステム 4 0 2 に相当するディレクトリツリーの最上位ディレクトリの下に、ディレクトリ「ba」及びディレクトリ「bb」がある。ディレクトリ「ba」の下には、ディレクトリ「bc」がある。

【 0 0 6 8 】

名前空間 4 0 0 A のルートディレクトリ「/」は、ファイルシステム 4 1 1 の最上位ディレクトリである。サーバ A は、ファイルシステム 4 1 1 に相当するディレクトリツリーにおけるパス「/mnt/fs0/」に、ファイルシステム 4 0 1 の最上位ディレクトリを繋ぎ、パス「/mnt/fs1/」に、ファイルシステム 4 0 2 の最上位ディレクトリを繋ぐことによって、一つの名前空間 4 0 0 A を構成している。このように、一つのサーバ 1 1 0 によって管理されるファイルシステムの間で、あるファイルシステムのディレクトリツリーに、他のファイルシステムのディレクトリツリーを繋ぐことを、ローカルマウントと呼ぶ。また、ファイルシステム 4 1 1 のように、ルートディレクトリを持つファイルシステムを、ルートファイルシステムと呼ぶ。

40

【 0 0 6 9 】

サーバ B は、二つのファイルシステム 4 1 2 及び 4 0 9 を管理する。それぞれのファイルシステムに属するディレクトリツリーを統合することによって、名前空間 4 0 0 B が構成される。

【 0 0 7 0 】

ファイルシステム 4 1 2 に相当するディレクトリツリーの最上位ディレクトリの下に、

50

ディレクトリ「mnt」がある。ディレクトリ「mnt」の下には、ディレクトリ「fs2」4 1 0がある。

【0071】

ファイルシステム409に相当するディレクトリツリーの最上位ディレクトリの下に、ディレクトリ「ca」がある。ディレクトリ「ca」の下に、ファイル「file2」及びファイル「file3」がある。

【0072】

名前空間400Bのルートファイルシステムは、ファイルシステム412である。ルートファイルシステム内のパス「/mnt/fs2」に、ファイルシステム409に相当するディレクトリツリーをローカルマウントすることによって、名前空間400Bが構成される。

10

【0073】

次に、以上で説明したサーバAの名前空間400Aと、サーバBの名前空間400Bの一部を統合し、共通名前空間であるGNS300を構成する手法を示す。

【0074】

名前空間400Aのパス「/gns/」をGNS300のルートディレクトリ「/」とする。GNS300のルートディレクトリを持つサーバAをGNSルートサーバと呼び、名前空間400Aのパス「/gns/」をGNSルートディレクトリと呼ぶ。GNSルートサーバとGNSルートディレクトリは、サーバソフトウェア200が記憶している。

【0075】

GNSルートディレクトリ（名前空間400Aの「/gns/」）以下のディレクトリツリーが、GNS300に属するディレクトリツリーとなる。GNSルートディレクトリの下には、接続ポイントを作ることができる。接続ポイントに、ファイルシステムに相当するディレクトリを繋げることによって、一つのディレクトリツリーを構成することができる。GNS300の接続ポイントにファイルシステムを繋げることによって一つのディレクトリツリーを構成することを、「接続」と呼ぶ。

20

【0076】

GNS300のGNSルートディレクトリの下には、三つの接続ポイント「/a/」403、「/b/」404及び「/c/」405がある。接続ポイント「/a/」403には、サーバAが管理するパス「/mnt/fs0/」にローカルマウントされているファイルシステム401が接続される。接続ポイント「/b/」404には、サーバAが管理するパス「/mnt/fs1/」にローカルマウントされているファイルシステム402が接続される。接続ポイント「/c/」405には、サーバBが管理するパス「/mnt/fs2/」にローカルマウントされているファイルシステム409が接続される。

30

【0077】

図5は、本発明の実施の形態のGNS管理表500の一例を示す説明図である。

【0078】

GNS管理表500は、各サーバ110のローカルメモリ113に格納され、各サーバ110によって個別に管理される。各サーバが格納するGNS管理表500には、同一の値が登録されている。GNS管理表500は、GNS300を構成する全てのファイルシステムを管理するためのテーブルであり、一つのエントリ（行）が、一つのディレクトリの接続ポイントを示す。

40

【0079】

サーバ名501には、ファイルシステムを管理するサーバ名が登録される。ローカルパス502には、サーバ名501に登録されたサーバ110が管理する名前空間において、ファイルシステムがローカルマウントされているパスが登録される。接続ポイント503は、ローカルパス502にローカルマウントしているディレクトリが、GNS300上のどの接続ポイントに接続されているかを示す。

【0080】

GNS管理表500は、さらに、GNSルートサーバ及びGNSルートディレクトリを示す情報を含んでもよい。具体的には、図3及び図4の例に対応するGNS管理表500

50

は、サーバAがGNSルートサーバであり、「/gns/」がGNSルートディレクトリであることを示す情報を含んでもよい。

【0081】

各サーバ110は、ルートファイルシステム内の名前解決と、そのルートファイルシステムに直接又は間接的にローカルマウントしているディレクトリ内の名前解決をすることができる。例えば、図3に示すGNS300の環境において、サーバ110AがGNS300上のファイル「/c/ca/file2」へのアクセス要求をクライアント150から受けたとする。サーバ110Aは、GNSルートサーバである。このため、サーバ110Aは、GNSルートディレクトリ「/gns/」をパス「/c/ca/file2」に付加し、GNS300のパスをローカルパス（サーバ110Aの名前空間におけるパス）に変換する。変換されたパスは、「/gns/c/ca/file2」となる。サーバAは、「/gns/」までの名前解決をすることはできる。しかし、サーバ110Aは、「/gns/c/」の名前解決をすることはできない。

10

【0082】

このとき、サーバ110Aは、GNS管理表500を参照し、「/gns/c/」がサーバ110Bの管理下にあることを判断する。そして、アクセス要求を発行したクライアント150に位置情報を通知する。ここで通知される位置情報は、そのディレクトリを管理しているサーバの識別子を含み、必要があれば、そのサーバ内のローカルパスを含んでもよい。通知を受けたクライアント150は、サーバ110Bにアクセス要求を発行することによって、GNS300上のファイル「/c/ca/file2」へアクセスすることができる。

【0083】

20

上記の例では、名前解決要求を受け取ったサーバAが、名前解決の対象となるファイルを管理しているサーバ110Bを識別する情報をクライアント150に通知する方法を示した。この方法とは異なり、名前解決要求を受け取ったサーバAがサーバBにクライアント150からの名前解決要求を転送し、サーバBが名前解決を実行してもよい。また、上記の例では、一つのファイルシステムが一つのサーバ110のみによって管理されていたが、複数のサーバ110が一つのファイルシステムを共有して管理してもよい。この場合、共有されたファイルシステムを更新する前に、更新を実行するサーバ110がファイルシステムのロックを取得することによって、更新が正しく実行されることを保証しなければならない。

【0084】

30

クライアント150にGNS300を提供する前に、管理計算機140は、予め、各サーバ110内にファイルシステムを分割する基点となるディレクトリパスを設定する。マイグレーションを実行する時に、このディレクトリパスを基点としてファイルシステムが分割され、分割されたファイルシステムがマイグレーションされる。その結果、よりきめ細かい粒度のマイグレーションが実現できる。このファイルシステムを分割する基点となるディレクトリパスをマイグレーションポイントと呼ぶ。各サーバ110は、それぞれ、マイグレーションポイント管理表600を使ってマイグレーションポイントを管理する。

【0085】

図6A及び図6Bは、本発明の実施の形態のマイグレーションポイント管理表600の一例を示す説明図である。

40

【0086】

図6Aにおけるマイグレーションポイント管理表600Aは、サーバAのローカルメモリ113Aに格納されるマイグレーションポイント管理表600である。図6Bにおけるマイグレーションポイント管理表600Bは、サーバBのローカルメモリ113Bに格納されるマイグレーションポイント管理表600である。以下において、マイグレーションポイント管理表600A及び600Bのいずれにも共通する説明をする場合、図6A及び図6Bに表示された符号から「A」、「B」を省略して、「マイグレーションポイント601」のように記載する。

【0087】

マイグレーションポイント601は、サーバ110内のローカルパスである。クライア

50

ント150がGNS300のファイルへアクセスするために使用されるGNSルートディレクトリと、GNS300を構成するために接続されているディレクトリとがマイグレーションポイント601として登録される。そして、各マイグレーションポイントにファイルシステムIDが与えられる。さらに、ファイルシステムを分割する基点としたいディレクトリもマイグレーションポイントとして登録され、予めファイルシステムIDが与えられる。

【0088】

例えば、図6Aのマイグレーションポイント管理表600Aには、マイグレーションポイント601として、GNS300のルートディレクトリ「/gns/」、GNS300に接続している「/mnt/fs0/」及び「/mnt/fs1/」、並びに、ファイルシステム分割の基点となる「/mnt/fs0/aa/」及び「/mnt/fs0/ab/ac/」が登録されている。

10

【0089】

システム管理者は、予め、任意のディレクトリをマイグレーションポイント601に登録することができる。

【0090】

ファイルシステムID602は、各ファイルシステム及び任意のディレクトリツリーに付与されたストレージシステム100内で一意の識別子である。管理計算機140は、マイグレーションポイント601をマイグレーションポイント管理表600に登録するとき、ストレージシステム100内で一意のファイルシステムIDを付与する。

20

【0091】

クライアント150が、マイグレーションポイント601より下位のファイル（すなわち、マイグレーションポイント601以下のディレクトリツリーに属するファイル）の名前解決を実行した場合、ファイルシステム処理プログラム202は、ファイルハンドル（識別子）をクライアントに返す。そのファイルハンドルは、アクセス対象のファイルと最も近い上位のディレクトリ階層にあるマイグレーションポイント601に対応するファイルシステムID902と、アクセスするファイルに付与されたファイルIDとを含む。例えば、GNSのパス「/a/aa/file1」の名前解決を実行した場合、そのGNSのパスは、サーバAのローカルパス「/mnt/fs0/aa/file1」に対応する（図3から図5参照）。「/mnt/fs0/aa/file1」に最も近い上位のマイグレーションポイント601は、「/mnt/fs0/aa/」である。このため、クライアントに返されるファイルハンドルは、マイグレーションポイント「/mnt/fs0/aa/」のファイルシステムID「2」、及び、ファイルID「file1」を含む。

30

【0092】

一度マイグレーションポイント601に与えたファイルシステムID602は、マイグレーションが実行された後も変える必要はない。そのため、ストレージシステム100内の各ファイルを識別するファイルハンドルは、マイグレーションポイント601以下のディレクトリツリー単位でマイグレーションが実行されても常に一定となる。この仕組みを利用して、クライアント150から透過的に（言い換えると、各ファイル又はディレクトリがどのサーバに管理されているかをクライアント150に意識させることなく）、マイグレーションポイントを基点としてファイルシステムを分割することが可能となる。詳細は後述する（図8及び図9参照）。

40

【0093】

使用容量603～writel/O情報605には、各マイグレーションポイントの統計情報が登録される。使用容量603は、マイグレーションポイント以下のディレクトリツリーが使用しているディスクの容量を示す。readl/O情報604には、ディレクトリツリーの中のファイルが何回readアクセスされたか、及び、ディレクトリツリーの中のファイルが何バイト読み込まれたかが登録される。writel/O情報605には、ディレクトリツリーの中のファイルが何回writeアクセスされたか、及び、ディレクトリツリーの中のファイルが何バイト書き込まれたかが登録される。

【0094】

50

例えば、readI/O情報 6 0 4 及びwriteI/O情報 6 0 5 は、毎秒の平均値でもよいし、所定の時間の累積値等であってもよい。さらに、使用容量 6 0 3 ~writeI/O情報 6 0 5 に登録される情報とは別に、例えば、毎分ごとにロギングした値が登録され、時間経過ごとの性能の変化が分かるようにしてもよい。

【 0 0 9 5 】

使用容量 6 0 3 ~writeI/O情報 6 0 5 に登録された情報は、どのサーバ 1 1 0 のマイグレーションポイントを、どここのサーバ 1 1 0 へマイグレーションするかを判断するための材料として使用することができる。詳細は後述する（図 1 6 及び図 1 9 参照）。

【 0 0 9 6 】

図 7 は、本発明の実施の形態のファイルシステム I D 管理表 7 0 0 の説明図である。

10

【 0 0 9 7 】

ファイルシステム I D 管理表 7 0 0 は、管理計算機 1 4 0 が参照することができるいずれかの場所に格納される。本実施の形態のファイルシステム I D 管理表 7 0 0 は、管理計算機 1 4 0 内のローカルメモリ 1 4 5 又はディスク 1 4 3 に格納される。

【 0 0 9 8 】

管理計算機 1 4 0 は、各マイグレーションポイントにストレージシステム 1 0 0 内で一意のファイルシステム I D を付与するために、ファイルシステム I D 管理表 7 0 0 を参照する。ファイルシステム I D 管理表 7 0 0 は、ファイルシステム I D 7 0 1 及び状態 7 0 2 からなる。

【 0 0 9 9 】

ファイルシステム I D 7 0 1 には、各マイグレーションポイントに既に付与された、又はこれから付与されるファイルシステム I D が登録される。

20

【 0 1 0 0 】

状態 7 0 2 には、各ファイルシステム I D 7 0 1 が既にいずれかのマイグレーションポイントに付与されているか否かを示す値が登録される。状態 7 0 2 に「済」が登録されたファイルシステム I D 7 0 1 は、既にストレージシステム 1 0 0 内でいずれかのマイグレーションポイントに付与されている。一方、状態 7 0 2 に「未」が登録されたファイルシステム I D 7 0 1 は、ストレージシステム 1 0 0 内でまだいずれのマイグレーションポイントにも付与されていない。

【 0 1 0 1 】

新たにマイグレーションポイントを設定するときに、管理計算機 1 4 0 は、そのマイグレーションポイントに、状態 7 0 2 の値が「未」であるファイルシステム I D 7 0 1 を付与する。そして、そのファイルシステム I D 7 0 1 に対応する状態 7 0 2 を「済」に更新する。その結果、新たに設定されるマイグレーションポイントには、常に、ストレージシステム内でまだどのマイグレーションポイントにも付与されていないファイルシステム I D 7 0 1 が付与される。

30

【 0 1 0 2 】

次に、本実施の形態において実行されるマイグレーション処理について説明する。

【 0 1 0 3 】

サーバ 1 1 0 にマイグレーションの実行契機を与えるいくつかの方法が考えられる。例えば、システム管理者が管理計算機 1 4 0 を介してサーバソフトウェア 2 0 0 にマイグレーションの指示を与えてもよい。あるいは、サーバソフトウェア 2 0 0 が自動的に判断してマイグレーション処理を開始してもよい。あるいは、別の方法によって実行契機が与えられてもよい。いずれの方法であっても本発明を実施することができる。

40

【 0 1 0 4 】

次に、本発明の実施例におけるマイグレーションの実行処理の手順を、図 8 及び図 9 を参照して説明する。なお、サーバソフトウェア 2 0 0 に含まれるプログラム及びモジュールは、ローカルメモリ 1 1 3 に格納され、C P U 1 1 2 によって実行される。このため、以下の説明において各プログラム又はモジュールが実行する処理は、実際には、C P U 1 1 2 によって実行される。

50

【 0 1 0 5 】

図 8 は、本発明の実施の形態のマイグレーション処理プログラム 2 0 7 が実行する処理のフローチャートである。

【 0 1 0 6 】

サーバ 1 1 0 が、マイグレーションの対象となるマイグレーションポイントを指定したマイグレーションの指示を管理計算機 1 4 0 から受信したとき、マイグレーション処理プログラム 2 0 7 は、図 8 に示す処理を開始する。あるいは、マイグレーション処理プログラム 2 0 7 自身が、例えば、マイグレーションポイント管理表 6 0 0 に登録された統計情報等を参照して、マイグレーションを実行するか否かを判定してもよい。マイグレーション処理プログラム 2 0 7 は、マイグレーションを実行すると判定した場合に、図 8 の処理を開始する。

10

【 0 1 0 7 】

最初に、マイグレーション処理プログラム 2 0 7 は、ファイルシステムを分割する必要があるか否かを判断する (8 0 1) 。

【 0 1 0 8 】

マイグレーション処理プログラム 2 0 7 が受信したマイグレーションポイント以下のディレクトリツリーがファイルシステム全体である場合 (言い換えると、受信したマイグレーションポイントがファイルシステムの最上位ディレクトリである場合)、ファイルシステム全体がマイグレーションされる。この場合、ファイルシステムを分割する必要がないため、判定の結果は「いいえ」となり、ステップ 8 0 4 の処理に進む。

20

【 0 1 0 9 】

一方、マイグレーション処理プログラム 2 0 7 が受信したマイグレーションポイント以下のディレクトリツリーがファイルシステムの一部である場合、ファイルシステムを分割する必要がある。このため、判定の結果は「はい」となり、ステップ 8 0 2 の処理に進む。

【 0 1 1 0 】

マイグレーション処理プログラム 2 0 7 は、ステップ 8 0 2 において、ファイルシステム分割処理モジュール 2 1 4 を呼び出す。ファイルシステム分割処理モジュール 2 1 4 が実行する処理の詳細については後述する (図 9 参照) 。

【 0 1 1 1 】

次に、マイグレーション処理プログラム 2 0 7 は、ファイルシステム分割処理 8 0 2 による分割の結果新しく作成されたファイルシステムをマイグレーションする必要があるか否かを判定する (8 0 3) 。

30

【 0 1 1 2 】

後述するように、ステップ 8 0 2 において局所分割が実行された場合、分割されたファイルシステムをマイグレーションする必要がある。一方、遠隔分割が実行された場合、分割されたファイルシステムをマイグレーションする必要がある。

【 0 1 1 3 】

ステップ 8 0 3 において、マイグレーションを実行する必要があると判定された場合、マイグレーション処理プログラム 2 0 7 は、分割されたファイルシステムのマイグレーションを実行して (8 0 4)、処理を終了する。

40

【 0 1 1 4 】

具体的には、マイグレーション処理プログラム 2 0 7 は、ステップ 8 0 4 において、マイグレーションの対象となる分割されたファイルシステムに含まれる全てのファイルを、マイグレーション先のサーバ 1 1 0 に送信する。マイグレーション先のサーバのマイグレーション処理プログラム 2 0 7 は、受信したファイルを、いずれかの L D E V 1 2 4 に格納する。

【 0 1 1 5 】

一方、ステップ 8 0 3 において、マイグレーションを実行する必要があると判定された場合、マイグレーション処理プログラム 2 0 7 は、マイグレーションを実行せずに、処理

50

を終了する。

【0116】

図9は、本発明の実施の形態のファイルシステム処理プログラム202内のファイルシステム分割処理モジュール214が実行する処理のフローチャートである。

【0117】

図9の処理は、図8のステップ802においてマイグレーション処理プログラム207によって呼び出されたときに、ファイルシステム分割処理モジュール214によって実行される。

【0118】

最初に、ファイルシステム分割処理モジュール214は、新しいファイルシステムを作成する(901)。ここで、ファイルシステム分割処理モジュール214は、自サーバ(すなわち、分割の対象のファイルシステムを管理しているサーバ110)が管理する新たなファイルシステムを作成してもよいし、自サーバ以外のサーバ110が管理する新たなファイルシステムを作成してもよい。前者のようにして実行される分割を、以下の説明において「局所分割」と記載する。後者のようにして実行される分割を、以下の説明において「遠隔分割」と記載する。局所分割の具体例は、後で図10から図12を参照して説明する。遠隔分割の具体例は、後で図13から図15を参照して説明する。

【0119】

次に、ファイルシステム分割処理モジュール214は、ステップ901において作成されたファイルシステムに、分割の対象となるマイグレーションポイント以下のディレクトリに属するファイルを、ファイルIDを維持したままコピーする(902)。マイグレーションによってファイルハンドルが変更されることを防ぐために、ファイルIDを維持する必要があるためである。このため、サーバ110のオペレーティングシステム(OS)は、ファイルIDを指定したファイルのコピーをする機能を備える必要がある。

【0120】

次に、ファイルシステム分割処理モジュール214は、マイグレーションポイント管理表600を更新する(903)。この更新については後述する。

【0121】

次に、ファイルシステム分割処理モジュール214は、GNS管理表500を更新する(904)。具体的には、ファイルシステム分割処理モジュール214は、GNS管理処理プログラム204を呼び出し、GNS管理表500を更新させる。この更新については後述する(図12及び図15参照)。

【0122】

次に、ファイルシステム分割処理モジュール214は、作成された新しいファイルシステムをGNSに接続する(905)。具体的には、マウントテーブル(図示省略)が更新される。マウントテーブルの更新は、従来と同様の方法によって実行されてもよい。このため、詳細な説明は省略する。

【0123】

以上で、ファイルシステム分割処理モジュール214は、処理を終了する。

【0124】

次に、ステップ901において自サーバが管理する新たなファイルシステムが作成される場合の図9の処理の具体例を、図10から図12を参照して説明する。

【0125】

具体的には、サーバAのマイグレーションポイント「/mnt/fs0/ab/ac/」407以下のディレクトリツリーを分割し、サーバAが管理する新しいファイルシステムを作成する場合を例として説明する。

【0126】

図10は、本発明の実施の形態において局所分割が実行されているときの名前空間の一例の説明図である。

【0127】

図 10 のうち、図 4 において既に説明した部分については、説明を省略する。

【0128】

ファイルシステム分割処理モジュール 214 は、サーバ A が利用可能な未使用の LDEV 124 に新しいファイルシステム 1002 を作成する (901)。さらに、ファイルシステム分割処理モジュール 214 は、作成された新しいファイルシステム 1002 をサーバ A のローカルパス「/mnt/fs3/」にローカルマウントする。

【0129】

次に、ファイルシステム分割処理モジュール 214 は、マイグレーションポイント 407 以下のディレクトリツリー 1001 に属するファイル「ad」及び「ae」を、ファイル ID を維持したまま、ファイルシステム 1002 以下のディレクトリツリーにコピーする (902)。ディレクトリツリー 1001 は、ファイルシステム 401 の一部に相当する。

【0130】

次に、ファイルシステム分割処理モジュール 214 は、マイグレーションポイント管理表 600A のマイグレーションポイント 601A 「/mnt/fs0/ab/ac/」を「/mnt/fs3/」に書き換える (903)。

【0131】

次に、GNS 管理処理プログラム 204 が GNS 管理表 500 を更新する (904)。具体的には、GNS 管理処理プログラム 204 は、GNS 管理表 500 に新たな行 1201 を追加する。更新された GNS 管理表 1200 は、図 12 に示す通りである。追加される行 1201 のサーバ名 501 には「サーバ A」、ローカルパス 502 には「/mnt/fs3/」、接続ポイント 503 には「/gns/a/ab/ac/」が登録される。

【0132】

次に、ファイルシステム分割処理モジュール 214 は、書き換えられた GNS 管理表 1200 に従って、サーバ A のローカルパス「/mnt/fs3/」1003 を接続ポイント「/gns/a/ab/ac/」1101 (図 11 参照) に接続する (905)。

【0133】

以上で、ファイルシステムの局所分割が終了する。

【0134】

図 11 は、本発明の実施の形態において局所分割が実行された後の名前空間の一例の説明図である。

【0135】

図 11 は、図 10 に示す局所分割の実行が終了した後の名前空間である。図 11 のうち、図 10 において既に説明した部分については、説明を省略する。

【0136】

サーバ A のローカルパス「/mnt/fs0/ab/ac/」の下ファイル「ad」及び「ae」が削除され、「/gns/a/ab/ac/」1101 が新たな接続ポイントとして設定されている。

【0137】

図 12 は、本発明の実施の形態において局所分割が実行された後の GNS 管理表 1200 の一例の説明図である。

【0138】

GNS 管理表 1200 には、図 10 において説明したように、新たな行 1201 が追加されている。

【0139】

局所分割が終了した後、新たに作成されたファイルシステムを他のサーバ 110 へマイグレーションする必要がある。図 11 の例では、ファイルシステム 1002 を、例えばサーバ B にマイグレーションする必要がある。このため、図 8 のステップ 803 において「はい」と判定され、ステップ 804 のマイグレーションが実行される。

【0140】

例えば、後述する図 13 に示すように、ファイルシステム 1002 がサーバ B にマイグレーションされる場合、マイグレーションが実行された後のサーバ A 及びサーバ B のロー

10

20

30

40

50

カル名前空間は、後述する図 1 4 に示す通りとなる。このとき、マイグレーションポイント管理表 6 0 0 A の「/mnt/fs3/」に対応する行は、マイグレーションポイント管理表 6 0 0 B に移動する。そして、マイグレーションが実行された後の G N S 管理表 5 0 0 は、後述する図 1 5 に示す通りとなる。

【 0 1 4 1 】

マイグレーションが実行される前、マイグレーションポイント管理表 6 0 0 A の「/mnt/fs3/」に対応するファイルシステム I D 6 0 2 A の値は「 3 」であった。このため、「/mnt/fs3/」に対応する行がマイグレーションポイント管理表 6 0 0 B に移動するときに、パス名「/mnt/fs3/」だけでなく、移動する行のファイルシステム I D 6 0 2 A の値「 3 」が、移動先のサーバ B に送信される。この送信は、L A N 1 6 0、サーバ間通信路 1 3 5 又はその他のネットワークを介して実行されてもよい。移動先のサーバ B は、移動元のサーバ A から送信されたパス名「/mnt/fs3/」及びファイルシステム I D 「 3 」を、それぞれ、マイグレーションポイント 6 0 1 B 及びファイルシステム I D 6 0 2 B に登録する。その結果、マイグレーションされたファイルは、マイグレーション実行前と同じファイルハンドルによって識別される。

【 0 1 4 2 】

ここで、ファイル I D を維持したままコピーする方法（図 9 のステップ 9 0 2 参照）について補足説明する。ファイル I D とは、各ファイルごとに付与される識別子であり、ファイルシステム内で一意となる。ファイル I D を維持したままコピーするためには、コピー先のファイルシステムに同じファイル I D を持つファイルがないことを保証しなければならない。しかし、ステップ 9 0 1 で作成されたばかりのファイルシステムには、登録されているファイルが無い場合、ファイル I D が重なることがないことが保証できる。そのため、ファイル I D を維持したままファイルをコピーすることが可能となる。言い換えると、一つのファイルがマイグレーションされたとき、マイグレーション実行後のマイグレーション先のサーバ 1 1 0 は、マイグレーション元のサーバ 1 1 0 がそのファイルを識別するために使用したファイル I D と同一のファイル I D を使用して、マイグレーションの対象のファイルを識別する。

【 0 1 4 3 】

クライアント 1 5 0 は、G N S のファイルにアクセスする場合、システム内で一意なファイルシステム I D とファイルシステム内で一意なファイル I D とを組み合わせて作成されるファイルハンドルを使用し、システム内のファイルを一意に識別する。サーバ 1 1 0 は、ファイルハンドルを指定したアクセス要求を受信すると、指定されたファイルハンドルによって識別されるファイルに対するアクセスを実行する。

【 0 1 4 4 】

本発明によれば、ファイルシステムの分割が実行される前後で、ファイルシステム I D もファイル I D も変わらない。また、ファイルシステム分割後にファイルシステムをマイグレーションしても、ファイルシステム I D とファイル I D のいずれも変わることはない。そのため、クライアント 1 5 0 は、ファイルシステム分割前に取得したファイルハンドルを使用して、ファイルシステム分割後も、ファイルシステムのマイグレーション後も、一意のファイルにアクセスすることができる。

【 0 1 4 5 】

分割によって作成される新しいファイルシステムを、あらかじめマイグレーション先のサーバ 1 1 0 に作成してもよい。そのようにして実行される分割が、遠隔分割である。例えば、分割元のファイルシステム 4 0 1 のマイグレーションポイント 4 0 7 以下のディレクトリツリーを分割し、新しいファイルシステムをサーバ B が管理するファイルシステムとして作成する場合に実行される図 9 の処理を、図 1 3 ~ 図 1 5 を参照して説明する。

【 0 1 4 6 】

図 1 3 は、本発明の実施の形態において遠隔分割が実行されているときの名前空間の一例の説明図である。

【 0 1 4 7 】

図 13 のうち、図 4 又は図 10 において既に説明した部分については、説明を省略する。

【0148】

ファイルシステム分割処理モジュール 214 は、サーバ B が利用可能な未使用の LDEV 124 に新しいファイルシステム 1302 を作成する (901)。さらに、ファイルシステム分割処理モジュール 214 は、作成された新しいファイルシステム 1302 をサーバ B のローカルパス「/mnt/fs3/」にローカルマウントする。

【0149】

次に、ファイルシステム分割処理モジュール 214 は、マイグレーションポイント「/mnt/fs0/ab/ac/」407 以下のディレクトリツリー 1301 以下のファイル「ad」及び「ae」を、ファイル ID を維持したまま、ファイルシステム 1302 にコピーする (902)。このコピーは、LAN 160 を介して実行されてもよいし、他のネットワークを介して実行されてもよい。

【0150】

次に、サーバ A のファイルシステム分割処理モジュール 214 は、マイグレーションポイント管理表 600B に「/mnt/fs3/」をマイグレーションポイントとして追加する指示をサーバ B に送信する (903)。

【0151】

この指示には、マイグレーションポイント管理表 600B に新たに追加されるマイグレーションポイント ID「/mnt/fs3/」と、マイグレーションポイント管理表 600A においてその ID と対応して登録されていたファイルシステム ID 602A の値「3」が含まれる。その結果、マイグレーションポイント管理表 600B に新たに追加される行のマイグレーションポイント ID 601B 及びファイルシステム ID 602B は、それぞれ、「/mnt/fs3/」及び「3」となる。その結果、マイグレーションされたファイルは、マイグレーション実行前と同じファイルハンドルによって識別される。

【0152】

次に、ファイルシステム分割処理モジュール 214 は、GNS 管理処理プログラム 204 に、GNS 管理表 500 を更新する指示を送信する (904)。GNS 管理処理プログラム 204 は、GNS 管理表 500 に新たな行 1501 を追加する。更新された GNS 管理表 1500 は、図 15 に示す通りである。追加される行 1501 のサーバ名 501 には「サーバ B」、ローカルパス 502 には「/mnt/fs3/」、接続ポイント 503 には「/gns/a/ab/ac/」が登録される。

【0153】

次に、ファイルシステム分割処理モジュール 214 は、書き換えられた GNS 管理表 1500 に従って、サーバ B のローカルパス「/mnt/fs3/」1303 を接続ポイント「/gns/a/ab/ac/」1401 (図 14 参照) に接続する (905)。

【0154】

以上で、ファイルシステムの遠隔分割が完了する。

【0155】

図 14 は、本発明の実施の形態において遠隔分割が実行された後の名前空間の一例の説明図である。

【0156】

図 14 は、図 13 に示す遠隔分割の実行が終了した後の名前空間である。図 14 のうち、図 13 において既に説明した部分については、説明を省略する。

【0157】

サーバ A のローカルパス「/mnt/fs0/ab/ac/」の下ファイル「ad」及び「ae」が削除され、「/gns/a/ab/ac/」1401 が新たな接続ポイントとして設定されている。

【0158】

図 15 は、本発明の実施の形態において遠隔分割が実行された後の GNS 管理表 1500 の一例の説明図である。

【 0 1 5 9 】

G N S 管理表 1 5 0 0 には、図 1 3 において説明したように、新たな行 1 5 0 1 が追加されている。

【 0 1 6 0 】

以上の遠隔分割によれば、分割の終了と同時に他のサーバ 1 1 0 へのファイルシステムのマイグレーションも終了している。このため、図 8 のステップ 8 0 3 において「いいえ」と判定され、ステップ 8 0 4 のマイグレーションは実行されない。

【 0 1 6 1 】

ファイルシステムの作成 9 0 1、ファイルのコピー 9 0 2、データの二重書き、及び、G N S 管理処理プログラム 2 0 4 との通信は、L A N 1 6 0 を介して実行されてもよいし、サーバ間通信路 1 3 5 を介して実行されてもよい。

10

【 0 1 6 2 】

図 1 0 ~ 図 1 3 の例のように、元のファイルシステムを管理するサーバ 1 1 0 に新しいファイルシステムを作成してからマイグレーションを実行する方法を局所分割マイグレーションと呼ぶ。図 1 3 ~ 図 1 5 の例のように、予めマイグレーション先のサーバ 1 1 0 に新しいファイルシステムを作成する方法を遠隔分割マイグレーションと呼ぶ。

【 0 1 6 3 】

局所分割マイグレーション及び遠隔分割マイグレーションには、それぞれ相反する長所、短所がある。局所分割マイグレーションの長所は、スイッチ 1 3 0 の設定を変更して、サーバ 1 1 0 とディスクコントローラ 1 2 5 間のパスを切り換えるだけでファイルシステムのマイグレーションができる場合、サーバ 1 1 0 間のファイルコピー等によるネットワーク帯域の消費を節約できることである。例えば、サーバ A からサーバ B へのマイグレーションが実行されるとき、サーバ A とサーバ B が同じディスクサブシステム 1 2 0 を共有していれば、スイッチ 1 3 0 の設定を変えることで、マイグレーションの対象のファイルシステムが入っている L D E V 1 2 4 をサーバ A からサーバ B の管理下に移すことができる。

20

【 0 1 6 4 】

局所分割マイグレーションの短所は、マイグレーション元のディスクサブシステム 1 2 0 の空き容量が足りない場合、及び、L D E V 1 2 4 を使い切ったために新たなファイルシステムを作れない場合に実行できないことである。

30

【 0 1 6 5 】

遠隔分割マイグレーションの長所は、マイグレーション元のディスクサブシステム 1 2 0 の空き容量が足りない場合、又は、L D E V 1 2 4 が足りない場合であっても、マイグレーション先のディスクサブシステム 1 2 0 の空き容量又は L D E V 数に余裕があれば、マイグレーションが可能なことである。

【 0 1 6 6 】

遠隔分割マイグレーションの短所は、スイッチ 1 3 0 の設定を変更して、サーバ 1 1 0 とディスクコントローラ 1 2 5 間のパスを切り換えるだけでファイルシステムのマイグレーションが済む場合であっても、ネットワーク帯域を消費してしまうことである。

40

【 0 1 6 7 】

本発明では、以上の局所分割マイグレーションと遠隔分割マイグレーションを組み合わせる使用することによって、より効率の良いファイルシステムのマイグレーションを実行することができる。

【 0 1 6 8 】

次に、マイグレーションの実行契機を与える具体例として、管理者が管理計算機 1 4 0 からサーバ 1 1 0 へマイグレーションの実行を指示するためのグラフィカルユーザインターフェース (G U I) を図 1 6 及び図 1 7 に例示する。

【 0 1 6 9 】

図 1 6 は、本発明の実施の形態の管理計算機 1 4 0 に表示されるマイグレーションポイント選択ウィンドウの説明図である。

50

【 0 1 7 0 】

マイグレーションポイント選択ウィンドウ 1 6 0 0 は、管理計算機 1 4 0 の管理画面 1 4 2 に表示される。マイグレーションポイント選択ウィンドウ 1 6 0 0 には、マイグレーション元サーバリスト 1 6 5 0、マイグレーションポイントリスト 1 6 6 0 及びマイグレーション先サーバリスト 1 6 7 0 が表示される。

【 0 1 7 1 】

マイグレーション元サーバリスト 1 6 5 0 は、システム管理者がどのサーバ 1 1 0 のディレクトリツリーをマイグレーションするかを決定するために表示される。マイグレーション元サーバリスト 1 6 5 0 は、サーバ名 1 6 0 1、空き容量 1 6 0 2、空き L D E V 数 1 6 0 3、C P U 負荷 1 6 0 4 及びネットワーク帯域 1 6 0 5 を含む。

10

【 0 1 7 2 】

サーバ名 1 6 0 1 には、マイグレーション元となるサーバ 1 1 0 の候補が表示される。空き容量 1 6 0 2 には、サーバ名 1 6 0 1 が示すサーバ 1 1 0 が利用できるディスクサブシステム 1 2 0 の空き容量が表示される。空き L D E V 数 1 6 0 3 には、サーバ名 1 6 0 1 が示すサーバ 1 1 0 が利用できるディスクサブシステム 1 2 0 に設定された L D E V の数が表示される。C P U 負荷 1 6 0 4 には、サーバ名 1 6 0 1 が示すサーバ 1 1 0 が現在使用している C P U 1 1 2 の負荷の値（例えば、%）が表示される。ネットワーク帯域 1 6 0 5 には、サーバ名 1 6 0 1 が示すサーバ 1 1 0 が現在使用しているネットワーク帯域の量（例えば、bytes毎秒）が表示される。システム管理者は、エントリの選択ボタン 1 6 0 6 を操作することによって、サーバ名 1 6 0 1 を選択することができる。選択されたサーバ名 1 6 0 1 が示すサーバ 1 1 0 が、マイグレーション元のサーバ 1 1 0 となる。

20

【 0 1 7 3 】

なお、システム管理者による選択ボタン 1 6 0 6 の操作は、入力装置 1 4 1 によって実行される。例えば、入力装置 1 4 1 がマウスを含む場合、システム管理者は、選択ボタン 1 6 0 6 上でマウスクリックをすることによって操作してもよい。実行ボタン 1 6 3 1 等についても同様である。

【 0 1 7 4 】

システム管理者は、マイグレーション元として、例えば、空き容量 1 6 0 2 が少ないサーバ 1 1 0 を選択してもよいし、空き L D E V 数 1 6 0 3 が少ないサーバ 1 1 0 を選択してもよいし、C P U 負荷 1 6 0 4 が高いサーバ 1 1 0 を選択してもよいし、ネットワーク帯域 1 6 0 5 の使用量の高いサーバ 1 1 0 を選択してもよい。あるいは、列 1 6 0 2 ~ 1 6 0 5 までの各種情報を組み合わせてマイグレーション元のサーバ 1 1 0 を選択してもよい。

30

【 0 1 7 5 】

マイグレーション元サーバのマイグレーションポイントリスト 1 6 6 0 には、選択ボタン 1 6 0 6 によって選択されたサーバ 1 1 0 のマイグレーションポイント管理表 6 0 0 が表示される。マイグレーションポイントリスト 1 6 6 0 に含まれるマイグレーションポイント 1 6 1 1、使用容量 1 6 1 2、read I/O 情報 1 6 1 3 及び write I/O 情報 1 6 1 4 には、それぞれ、マイグレーションポイント管理表 6 0 0 に登録されているマイグレーションポイント 6 0 1、使用容量 6 0 3、read I/O 情報 6 0 4 及び write I/O 情報 6 0 5 の値が表示される。

40

【 0 1 7 6 】

システム管理者は、選択ボタン 1 6 1 5 を操作することによって、マイグレーションポイントリストに表示されたマイグレーションポイントの中から、マイグレーションしたいディレクトリツリーを選択する。具体的には、システム管理者は、マイグレーションしたいディレクトリツリーの最上位ディレクトリを、マイグレーションポイント 1 6 1 1 から選択する。

【 0 1 7 7 】

マイグレーション先サーバリスト 1 6 7 0 には、サーバ名 1 6 2 1、空き容量 1 6 2 2、空き L D E V 数 1 6 2 3、C P U 負荷 1 6 2 4 及びネットワーク帯域 1 6 2 5 が表示さ

50

れる。システム管理者は、選択ボタン 1 6 1 5 を操作して選択したマイグレーションポイント以下のディレクトリツリーをマイグレーションする先のサーバ 1 1 0 を、マイグレーション先サーバリスト 1 6 7 0 から選択する。この選択は、システム管理者が選択ボタン 1 6 2 6 を操作することによって実行される。

【 0 1 7 8 】

サーバ名 1 6 2 1 には、マイグレーション先の候補となるサーバ 1 1 0 の名前が表示される。空き容量 1 6 2 2 は、サーバ名 1 6 2 1 が示すサーバ 1 1 0 が利用できるディスクサブシステム 1 2 0 の空き容量が表示される。空き L D E V 数 1 6 2 3 には、サーバ名 1 6 2 1 が示すサーバ 1 1 0 が利用できるディスクサブシステムの空き L D E V の数が表示される。C P U 負荷 1 6 2 4 には、サーバ名 1 6 2 1 が示すサーバ 1 1 0 が現在使用している C P U 1 1 2 の負荷の値（例えば、％）が表示される。ネットワーク帯域 1 6 2 5 には、サーバ名 1 6 2 1 が示すサーバ 1 1 0 が現在使用しているネットワーク帯域の量（例えば、bytes毎秒）が表示される。

10

【 0 1 7 9 】

システム管理者がマイグレーション実行ボタン 1 6 3 1 を操作すると、マイグレーション先の詳細入力ウィンドウ 1 7 0 0 が表示される（図 1 7 参照）。キャンセルボタン 1 6 3 2 は、マイグレーションポイント選択ウィンドウ 1 6 0 0 を閉じ、マイグレーション処理をキャンセルするために使用される。

【 0 1 8 0 】

図 1 7 は、本発明の実施の形態のマイグレーション先詳細設定ウィンドウ 1 7 0 0 の説明図である。

20

【 0 1 8 1 】

マイグレーション先詳細設定ウィンドウ 1 7 0 0 は、マイグレーション元サーバ名表示欄 1 7 0 1、マイグレーションポイント表示欄 1 7 0 2、マイグレーション先サーバ名表示欄 1 7 0 3、L D E V 名入力欄 1 7 0 4、ローカルパス入力欄 1 7 0 5、実行ボタン 1 7 3 1 及びキャンセルボタン 1 7 3 2 を含む。

【 0 1 8 2 】

マイグレーション元サーバ名表示欄 1 7 0 1 には、図 1 6 の選択ボタン 1 6 0 6 によって選択されたマイグレーション元のサーバ 1 1 0 のサーバ名 1 6 0 1 が表示される。図 1 7 の例では、「サーバ A」が表示されている。マイグレーションポイント表示欄 1 7 0 2 には、選択ボタン 1 6 1 5 によって選択されたマイグレーションポイント 1 6 1 1 が表示される。図 1 7 の例では、「/mnt/fs3/」が表示されている。マイグレーション先サーバ名表示欄 1 7 0 3 には、選択ボタン 1 6 2 6 によって選択されたマイグレーション先のサーバ 1 1 0 のサーバ名 1 6 2 1 が表示される。図 1 7 の例では、「サーバ B」が表示されている。

30

【 0 1 8 3 】

システム管理者は、マイグレーション先のサーバ 1 1 0 が利用することができる L D E V 1 2 4 の中から、マイグレーション先として使用するものを選択し、その L D E V 1 2 4 の L D E V 名を L D E V 名入力欄 1 7 0 4 に入力する。

【 0 1 8 4 】

40

さらに、システム管理者は、マイグレーション先の L D E V 1 2 4 に作成されるファイルシステムをローカルマウントするマイグレーション先のサーバ 1 1 0 のローカルパスを指定し、そのローカルパスをローカルパス入力欄 1 7 0 5 に入力する。その結果、マイグレーション先サーバ名表示欄 1 7 0 3 に表示されたサーバ 1 1 0 の、ローカルパス入力欄 1 7 0 5 に入力されたローカルパスに、マイグレーション先のファイルシステムが作成される。ローカルパス入力欄 1 7 0 5 に入力される値は、G N S 管理表 5 0 0 のローカルパス 5 0 2 に対応している。

【 0 1 8 5 】

システム管理者が実行ボタン 1 7 3 1 を操作すると、管理計算機 1 4 0 は、サーバソフトウェア 2 0 0 へマイグレーションの開始を指示する。システム管理者がキャンセルボタ

50

ン 1 7 3 2 を選択すると、マイグレーション先詳細設定ウィンドウ 1 7 0 0 が閉じる。

【 0 1 8 6 】

局所分割マイグレーションと遠隔分割マイグレーションは、自動的に使い分けられてもよい。例えば、サーバ A からサーバ B へのマイグレーションが実行されるとき、サーバ A とサーバ B が同じディスクサブシステム 1 2 0 を共有している場合、L D E V 名入力欄 1 7 0 4 に入力された L D E V 1 2 4 へのマイグレーションは、パスを切り替えるだけで実行することができる。このとき、局所分割に使用する L D E V 1 2 4 のファイルシステムを一時的にサーバ A にローカルマウントするためのディレクトリは、管理計算機 1 4 0 によって自動生成されてもよい。

【 0 1 8 7 】

サーバ A とサーバ B が同じディスクサブシステム 1 2 0 を共有していない場合、パス切り替えによってマイグレーションを実行することができない。このため、自動的に、遠隔分割マイグレーションが選択され、実行される。その場合は、L D E V 名入力欄 1 7 0 4 に入力された L D E V 1 2 4 がサーバ B に確保され、その L D E V 1 2 4 にファイルシステムが作成される。そして、作成されたファイルシステムに、マイグレーションポイント表示欄 1 7 0 2 に表示されたディレクトリ以下のファイルを、ファイル I D を維持したままコピーすることによって、マイグレーションが実行される。

【 0 1 8 8 】

図 1 6 及び図 1 7 では、システム管理者が G U I ウィンドウを使用してマイグレーション開始の契機を与える方法について説明した。しかし、G U I ウィンドウは、マイグレーションの契機を与える方法のうちの一つであり、他の方法によってマイグレーションの契機が与えられてもよい。例えば、コマンドラインインターフェース (C L I) によってマイグレーションの契機が与えられてもよい。あるいは、マイグレーション元サーバの統計情報、マイグレーションポイント管理表の統計情報及びマイグレーション先サーバの統計情報に基づいて、どのサーバ 1 1 0 のどのマイグレーションポイントをどのサーバ 1 1 0 へマイグレーションするか自動的に判断してもよい。

【 0 1 8 9 】

図 1 8 は、本発明の実施の形態の管理計算機 1 4 0 が統計情報に基づいてマイグレーションの実行を制御する処理を説明するフローチャートである。

【 0 1 9 0 】

図 1 8 の例では、統計情報として、各サーバの空き容量 1 6 0 2、C P U 負荷 1 6 0 4 及びネットワーク帯域使用量 1 6 0 5 が使用される。具体的には、これらの値と、予め設定された閾値とを比較し、その比較の結果に基づいて、マイグレーションの実行が制御される。

【 0 1 9 1 】

なお、以下に説明するステップ 1 8 0 1 から 1 8 0 7 までの処理は、マイグレーション元サーバリスト 1 6 5 0 に表示された各サーバ 1 1 0 を対象として実行される。

【 0 1 9 2 】

管理計算機 1 4 0 の C P U 1 4 4 は、最初に、空き容量 1 6 0 2 の値が所定の閾値を下回っているか否かを判定する (1 8 0 1) 。

【 0 1 9 3 】

ステップ 1 8 0 1 において、閾値を下回っていると判定された場合 (すなわち、「はい」の場合)、特定の L D E V 1 2 4 にデータが集中して格納されていると考えられる。この場合、L D E V 1 2 4 に格納されているデータ量を分散させるために、マイグレーションを実行することが望ましい。この場合、処理はステップ 1 8 0 4 へ進む。

【 0 1 9 4 】

この場合、C P U 1 4 4 はステップ 1 8 0 4 において、空き容量 1 6 0 2 の閾値を上回るために十分な使用容量 1 6 1 2 を持つ一つ以上のマイグレーションポイントを選択する。

【 0 1 9 5 】

一方、ステップ1801において、閾値を下回っていないと判定された場合（すなわち、「いいえ」の場合）、処理はステップ1802へ進む。

【0196】

ステップ1802において、CPU144は、CPU負荷1604の値が所定の閾値を上回ったか否かを判定する。

【0197】

ステップ1802において、閾値を上回っていると判定された場合（すなわち、「はい」の場合）、特定のサーバに処理の負荷が集中していると考えられる。この場合、サーバ110のCPU負荷を分散させるために、マイグレーションを実行することが望ましい。この場合、処理はステップ1804へ進む。

10

【0198】

この場合、CPU144はステップ1804において、クライアント150から各マイグレーションポイントへのアクセスによるCPU負荷を、readI/O情報1613及びwriteI/O情報1614に基づいて算出する。正確なCPU負荷を算出する代わりに、近似的な値を算出してもよい。そして、CPU144は、マイグレーションによってCPU負荷1604が閾値を下回るように、一つ以上の適切なマイグレーションポイントを選択する。

【0199】

一方、ステップ1802において、閾値を上回っていないと判定された場合（すなわち、「いいえ」の場合）、処理はステップ1803へ進む。

【0200】

20

ステップ1803において、CPU144は、ネットワーク帯域1605の値（すなわち、ネットワーク帯域の使用量）が所定の閾値を上回ったか否かを判定する。

【0201】

ステップ1803において、閾値を上回ったと判定された場合（すなわち、「はいの場合」）、特定のサーバに通信の負荷が集中していると考えられる。この場合、通信の負荷を分散させるために、マイグレーションを実行することが望ましい。この場合、処理はステップ1804へ進む。

【0202】

この場合、CPU144はステップ1804において、readバイト数1613及びwriteバイト数1614に基づいて、各マイグレーションポイントの単位時間当たりのネットワーク帯域の使用量を計算する。そして、CPU144は、その計算の結果に基づいて、ネットワーク帯域1605が閾値を下回るように、一つ以上の適切なマイグレーションポイントを選択する。

30

【0203】

一方、ステップ1803において、閾値を上回っていないと判定された場合（すなわち、「いいえ」の場合）、データ量及び負荷のいずれも集中していない。この場合、マイグレーションを実行する必要がないため、処理を終了する。

【0204】

CPU144は、ステップ1804を実行した後、マイグレーションの条件に適合するサーバ110があるか否かを判定する（1805）。

40

【0205】

具体的には、CPU144は、ステップ1804で選択されたマイグレーションポイントの数以上の空きLDEV数1623を持つサーバ110の中から、ステップ1804で選択されたマイグレーションポイントをマイグレーションしたとしても、次の三つの条件を満たすと予測されるサーバ110を選択する。

（1）空き容量1602が所定の閾値を下回らない。

（2）CPU負荷1604が所定の閾値を上回らない。

（3）ネットワーク帯域1605が所定の閾値を上回らない。

【0206】

ステップ1805において、マイグレーションの条件に適合するサーバ110があると

50

判定された場合、そのサーバ110へのマイグレーションを実行することによって、サーバ110の負荷等が分散される。このため、CPU144は、条件に適合するサーバ110をマイグレーション先として選択する(1806)。複数のサーバ110が条件に適合する場合、それらのうちの任意の数のサーバ110がマイグレーション先として選択されてもよい。

【0207】

次に、CPU144は、ステップ1804において選択されたマイグレーションポイントを、ステップ1806において選択されたサーバ110へマイグレーションする指示を、サーバ110に送信する(1807)。複数のマイグレーションポイントをマイグレーションする場合、全てのマイグレーションポイントを一つのサーバ110へマイグレーションしてもよいし、複数のサーバへ分散してマイグレーションしてもよい。ただし、複数のサーバ110全体で、選択されたマイグレーションポイントの数の空きLDEV数1623を用意することができ、かつ、全てのサーバが上記の条件(1)から(3)を満たす必要がある。

10

【0208】

一方、ステップ1805において、マイグレーションの条件に適合するサーバ110がないと判定された場合、マイグレーションを実行してもサーバ110の負荷等が分散されない。この場合、CPU144は、処理を終了する。

【0209】

以上の処理を実行することによって、全サーバ110の空き容量1602、CPU負荷1604、ネットワーク帯域1605を適切な範囲に保つことができる。

20

【0210】

また、各統計情報の履歴を取っておくことによって、次のような処理も実現できる。

【0211】

例えば、各マイグレーションポイントの使用容量1612、readI/O情報1613及びwritel/O情報1614の履歴を参照することによって、ほとんど読み込みにしか使っていないマイグレーションポイントであるか否かが分かる。ほとんど読み込みにしか使っていないマイグレーションポイントは、予めアーカイブ用途に適したLDEV124を持つサーバ110へマイグレーションされてもよい。

【0212】

30

また、使用容量1612とwritel/O情報1614の履歴を参照することによって、著しく使用容量が増加傾向にあるマイグレーションポイントを見つけることができる。このようなマイグレーションポイントは、特に空き容量1602が豊富で、書き込み性能の良いLDEV124を持つサーバ110へ予めマイグレーションすることもできる。

【0213】

また、readI/Oバイト数1613及びwritel/Oバイト数1614の履歴を参照することによって、各マイグレーションポイントのネットワーク帯域の消費量が増加傾向にあるか否かが分かる。ネットワーク帯域の消費量が増加傾向にあるマイグレーションポイントは、特に大きいネットワーク帯域を持つサーバ110へ予めマイグレーションすることもできる。

40

【0214】

この他、短時間にマイグレーションを繰り返すことによってシステム全体の性能が低下することを避けるために、一度マイグレーションされたマイグレーションポイントは、しばらくマイグレーションを禁止することもできる。

【0215】

サーバ110の統計情報(すなわち、マイグレーション元サーバリスト1650及びマイグレーション先サーバリスト1670に表示される情報)は、例えば、管理計算機140がマイグレーションのウィンドウを表示するときに取得してもよいし、周期的に取得してもよい。サーバ110間の通信には、LAN160が使用されてもよいし、サーバ間通信路135が使用されてもよい。

50

【 0 2 1 6 】

次に、上記の本発明の実施の形態の効果について説明する。

【 0 2 1 7 】

クライアントから透過的に、きめ細かいマイグレーションを実現する従来の方法として、予めストレージシステム 1 0 0 に少ない容量のファイルシステムを多数用意するという方法があった。以下、図 1 9 を参照して、本実施の形態と上記従来の方法とを比較する。

【 0 2 1 8 】

図 1 9 は、本発明の実施の形態と従来の手法とを比較する説明図である。

【 0 2 1 9 】

説明の便宜上、1 0 0 G B の記憶容量を持つストレージシステム 1 9 0 0 A 及び 1 9 0 0 B があると仮定する。ストレージシステム 1 9 0 0 A には本発明の実施の形態が適用され、ストレージシステム 1 9 0 0 B には従来の方法が適用される。ストレージシステム 1 9 0 0 A は、一つのファイルシステム 1 9 0 1 A のみを含む。すなわち、ファイルシステム 1 9 0 1 A の記憶容量は 1 0 0 G B である。一方、ストレージシステム 1 9 0 0 B には、予め小さいファイルシステムが用意される。具体的には、ストレージシステム 1 9 0 0 B は、各々の記憶容量が 2 5 G B である四つのファイルシステム 1 9 0 1 B、1 9 0 2 B、1 9 0 3 B 及び 1 9 0 4 B を含む。

【 0 2 2 0 】

ストレージシステム 1 9 0 0 A 及び 1 9 0 0 B に格納されているファイルのデータ量は、それぞれ 5 0 G B である。図 1 9 中の網掛けされている領域が、ストレージシステム 1 9 0 0 A 及び 1 9 0 0 B の使用領域（すなわち、ファイルを格納している領域）である。ストレージシステム 1 9 0 0 B では、ファイルシステム 1 9 0 1 B 及び 1 9 0 2 B にそれぞれ 2 0 G B のデータが格納され、ファイルシステム 1 9 0 3 B には 1 0 G B のデータが格納され、ファイルシステム 1 9 0 4 B にはデータが格納されていない。

【 0 2 2 1 】

図 1 9 の例では、各ファイルシステムの使用率が 8 0 % になった時点で、マイグレーションが実行されると仮定する。ファイルシステムの使用率とは、各ファイルシステムに設定されている記憶容量に対する、実際に使用されている容量の割合（%）である。

【 0 2 2 2 】

以上の条件で、図 1 9 を参照すると、ストレージシステム 1 9 0 0 A では、ファイルシステム 1 9 0 1 A の使用率が 5 0 % である。このため、マイグレーションを実行する必要が無い。これに対し、ストレージシステム 1 9 0 0 B では、ファイルシステム 1 9 0 1 B 及び 1 9 0 2 B の使用率は、マイグレーションの閾値である 8 0 % である。このため、ファイルシステム 1 9 0 1 B 及び 1 9 0 2 B は、マイグレーションの対象となってしまう。このように、ファイルシステム 1 9 0 1 B 及び 1 9 0 2 B に格納されるファイルが集中している一方で、ファイルシステム 1 9 0 3 B 及び 1 9 0 4 B には、空き容量の余裕がある。ストレージシステム 1 9 0 0 B では、マイグレーションの回数が増える上に、最初に割り当てられたファイルシステム 1 9 0 4 B が無駄になる可能性もあり、性能面においても、ストレージシステム 1 0 0 の使用効率においても無駄が多い。

【 0 2 2 3 】

小さなファイルシステムを予め作成し、細かい粒度のマイグレーションを実現する従来の方法で、効率よくストレージシステム 1 0 0 を使用するためには、図 1 9 の例のようなことが起こらないように、予めファイルシステムごとの細かい運用計画を立てなければならない。しかしながら、実際にストレージシステム 1 0 0 の運用の前から細かい予測を立てることは難しい。

【 0 2 2 4 】

これに対して、本発明の実施の形態によれば、必要に応じてファイルシステムを分割してマイグレーションを実行することによって、ファイルシステム数を必要最小限に抑えることができる。このため、本実施の形態は、図 1 9 のストレージシステム 1 9 0 0 B に示す方法と比べ、マイグレーション処理の増加による性能劣化、及び、容量効率の無駄を抑

10

20

30

40

50

える効果がある。

【 0 2 2 5 】

以上、本発明の実施の形態によれば、複数のサーバ 1 1 0 を備えるストレージシステム 1 0 0 において、ファイルシステムを分割することによって、きめ細かい粒度のマイグレーションを、クライアントから透過的に実現することが可能となる。これは、複数のサーバ 1 1 0 間のより良い負荷の均衡を可能とし、ストレージシステム 1 0 0 全体の性能向上に寄与する。

【図面の簡単な説明】

【 0 2 2 6 】

【図 1】本発明の実施の形態のストレージシステムの構成を示すブロック図である。

10

【図 2】本発明の実施の形態のサーバソフトウェアの説明図である。

【図 3】本発明の実施の形態においてクライアントに提供される共通名前空間の一例を示す説明図である。

【図 4】本発明の実施の形態の各サーバの名前空間の説明図である。

【図 5】本発明の実施の形態の G N S 管理表の一例を示す説明図である。

【図 6 A】本発明の実施の形態のマイグレーションポイント管理表の一例を示す説明図である。

【図 6 B】本発明の実施の形態のマイグレーションポイント管理表の一例を示す説明図である。

【図 7】本発明の実施の形態のファイルシステム I D 管理表の説明図である。

20

【図 8】本発明の実施の形態のマイグレーション処理プログラムが実行する処理のフローチャートである。

【図 9】本発明の実施の形態のファイルシステム処理プログラム内のファイルシステム分割処理モジュールが実行する処理のフローチャートである。

【図 1 0】本発明の実施の形態において局所分割が実行されているときの名前空間の一例の説明図である。

【図 1 1】本発明の実施の形態において局所分割が実行された後の名前空間の一例の説明図である。

【図 1 2】本発明の実施の形態において局所分割が実行された後の G N S 管理表の一例の説明図である。

30

【図 1 3】本発明の実施の形態において遠隔分割が実行されているときの名前空間の一例の説明図である。

【図 1 4】本発明の実施の形態において遠隔分割が実行された後の名前空間の一例の説明図である。

【図 1 5】本発明の実施の形態において遠隔分割が実行された後の G N S 管理表の一例の説明図である。

【図 1 6】本発明の実施の形態の管理計算機に表示されるマイグレーションポイント選択ウィンドウの説明図である。

【図 1 7】本発明の実施の形態のマイグレーション先詳細設定ウィンドウの説明図である。

40

【図 1 8】本発明の実施の形態の管理計算機が統計情報に基づいてマイグレーションの実行を制御する処理を説明するフローチャートである。

【図 1 9】本発明の実施の形態と従来の手法とを比較する説明図である。

【符号の説明】

【 0 2 2 7 】

1 0 0 ストレージシステム

1 1 0 サーバ

1 1 1 ネットワークインターフェース

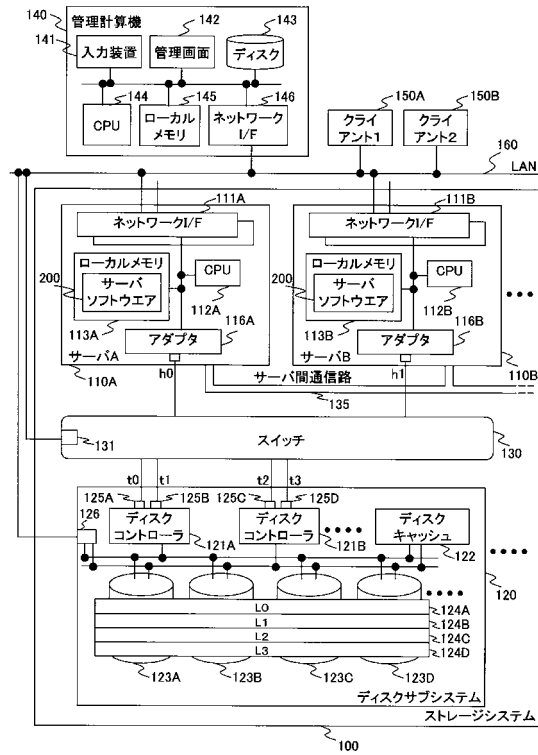
1 1 2 C P U

1 1 3 ローカルメモリ

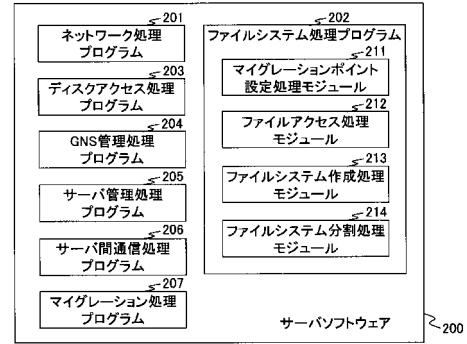
50

1 1 6	アダプタ	
1 2 0	ディスクサブシステム	
1 2 1	ディスクコントローラ	
1 2 2	ディスクキャッシュ	
1 2 3	ディスクドライブ	
1 2 4	論理デバイス (L D E V)	
1 2 5	ポート	
1 2 6	管理用ポート	
1 3 0	スイッチ	
1 3 1	管理用ポート	10
1 3 5	サーバ間通信路	
1 4 0	管理計算機	
1 5 0	クライアント	
1 6 0	L A N	
2 0 0	サーバソフトウェア	
2 0 1	ネットワーク処理プログラム	
2 0 2	ファイルシステム処理プログラム	
2 0 3	ディスクアクセス処理プログラム	
2 0 4	G N S 管理処理プログラム	
2 0 5	サーバ管理処理プログラム	20
2 0 6	サーバ間通信処理プログラム	
2 0 7	マイグレーション処理プログラム	
2 1 1	マイグレーションポイント設定処理モジュール	
2 1 2	ファイルアクセス処理モジュール	
2 1 3	ファイルシステム作成処理モジュール	
2 1 4	ファイルシステム分割処理モジュール	

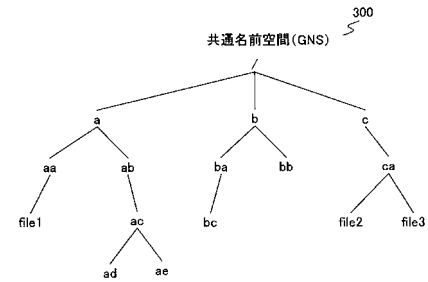
【図 1】



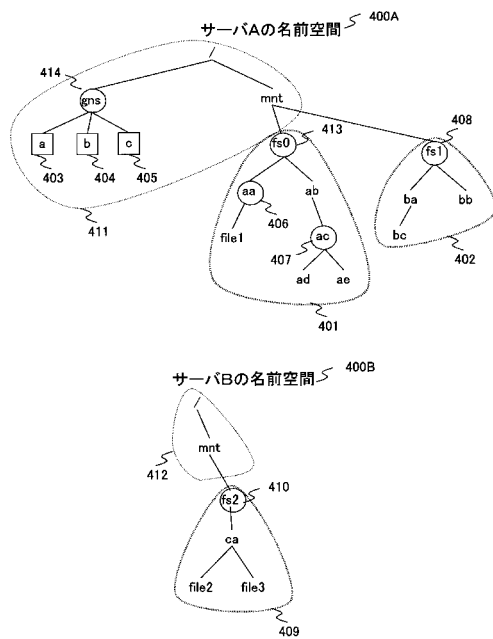
【図 2】



【図 3】



【図 4】



【図 5】

GNS管理表

サーバ名	ローカルパス	接続ポイント
サーバA	/mnt/fs0/	/gns/a/
サーバA	/mnt/fs1/	/gns/b/
サーバB	/mnt/fs2/	/gns/c/

【図 6 A】

マイグレーションポイント管理表(サーバA)

マイグレーションポイント	ファイルシステムID	使用容量	read I/O情報 回数	read I/O情報 バイト数	write I/O情報 回数	write I/O情報 バイト数
/gns/	0					
/mnt/fs0/	1					
/mnt/fs0/aa/	2					
/mnt/fs0/ab/ac/	3					
/mnt/fs1/	4					

【図 6 B】

マイグレーションポイント管理表(サーバB)

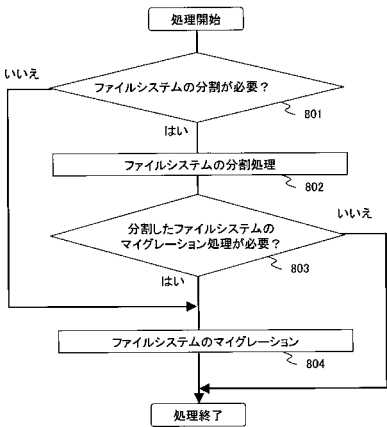
601B マイグレーション ポイント	602B ファイルシステム ID	603B 使用 容量	604B read I/O情報 回数 バイト数	605B write I/O情報 回数 バイト数
/mnt/fs2/	6			

【図 7】

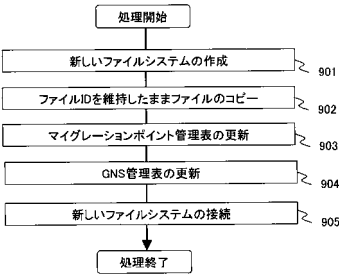
ファイルシステムID管理表

701 ファイルシステムID	702 状態
0	済
1	済
2	済
3	済
4	済
5	未
6	済
7	未

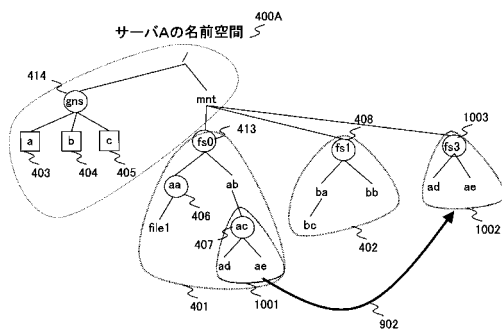
【図 8】



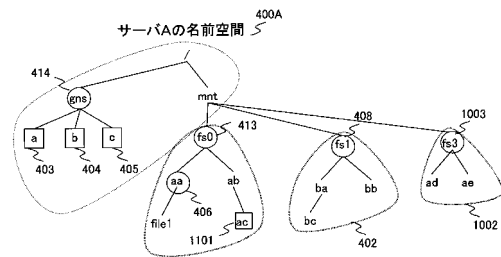
【図 9】



【図 10】



【図 11】

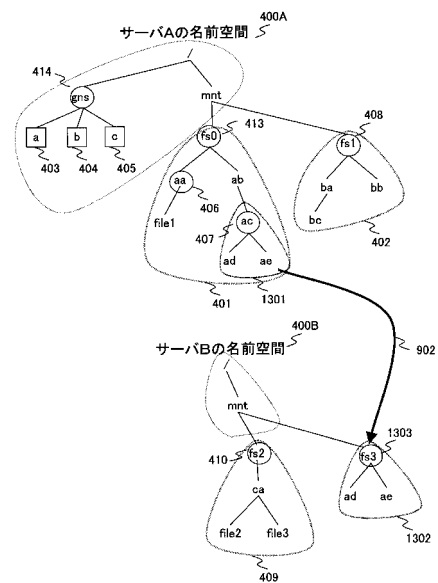


【図 12】

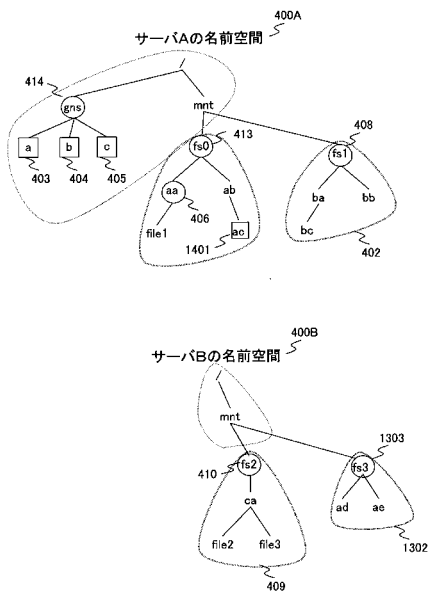
GNS管理表

サーバ名	ローカルパス	接続ポイント
サーバA	/mnt/fs0/	/gns/a/
サーバA	/mnt/fs1/	/gns/b/
サーバB	/mnt/fs2/	/gns/c/
サーバA	/mnt/fs3/	/gns/a/ab/ac/

【図 13】



【図 14】



【図 15】

GNS管理表		
サーバ名	ローカルパス	接続ポイント
サーバA	/mnt/fs0/	/gns/a/
サーバA	/mnt/fs1/	/gns/b/
サーバB	/mnt/fs2/	/gns/c/
サーバB	/mnt/fs3/	/gns/a/ab/ac/

【図 16】

1601

1602

1603

1604

1605

1650

サーバ名	空き容量	空きLDEV数	CPU負荷	ネットワーク帯域
サーバA				
サーバB				

●

マイグレーションポイントリスト

1606

1611

1612

1613

1614

1660

マイグレーションポイント	使用容量	read/O情報		write/O情報	
		回数	バイト数	回数	バイト数
/gns/					
/mnt/fs0/					
/mnt/fs0/aa/					
/mnt/fs3/					
/mnt/fs1/					

○

○

○

●

○

1615

マイグレーション先サーバリスト

1621

1622

1623

1624

1625

1670

サーバ名	空き容量	空きLDEV数	CPU負荷	ネットワーク帯域
サーバA				
サーバB				

○

●

1626

実行

キャンセル

1631

1632

1600

【図 17】

マイグレーション先詳細設定

マイグレーション元サーバ名:

サーバA

マイグレーションポイント:

/mnt/fs3/

マイグレーション先サーバ名:

サーバB

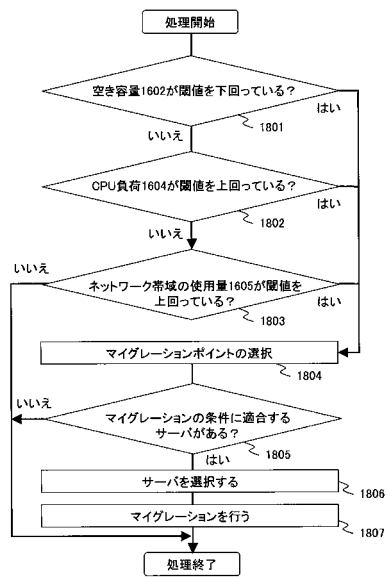
LDEV名:

ローカルパス:

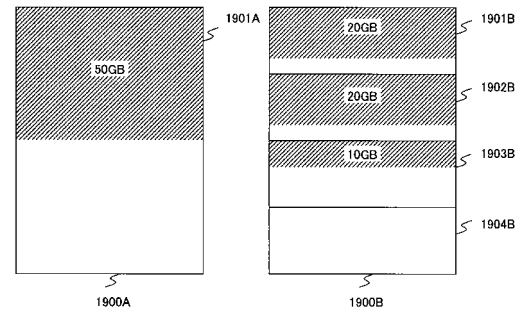
実行

キャンセル

【図 18】



【図 19】



フロントページの続き

(72)発明者 中谷 洋司

神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所 システム開発研究所内

(72)発明者 中村 隆喜

神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所 システム開発研究所内

審査官 桜井 茂行

(56)参考文献 特開 2 0 0 6 - 0 9 2 3 2 2 (J P , A)

(58)調査した分野(Int.Cl. , D B 名)

G 0 6 F 1 2 / 0 0

G 0 6 F 3 / 0 6

G 0 6 F 1 3 / 1 0