

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4291077号
(P4291077)

(45) 発行日 平成21年7月8日(2009.7.8)

(24) 登録日 平成21年4月10日(2009.4.10)

(51) Int.Cl. F I
G 0 6 F 12/00 (2006.01)
 G 0 6 F 12/00 5 4 5 A
 G 0 6 F 12/00 5 0 1 A
 G 0 6 F 12/00 5 3 1 D

請求項の数 8 (全 20 頁)

<p>(21) 出願番号 特願2003-282049 (P2003-282049) (22) 出願日 平成15年7月29日 (2003.7.29) (65) 公開番号 特開2005-50165 (P2005-50165A) (43) 公開日 平成17年2月24日 (2005.2.24) 審査請求日 平成18年3月22日 (2006.3.22)</p>	<p>(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 100084537 弁理士 松田 嘉夫 (74) 代理人 100114236 弁理士 藤井 正弘 (72) 発明者 坂口明彦 東京都国分寺市東恋ヶ窪一丁目280番地 株式会社日立製作所 中央研究所内</p>
--	---

最終頁に続く

(54) 【発明の名称】 分散ストレージ装置のファイル管理方法及び分散ストレージシステム

(57) 【特許請求の範囲】

【請求項1】

分散配置された複数のストレージ装置と、

これらストレージ装置が格納するファイルの識別子及びストレージ装置を管理する管理サーバとを備えて、前記分散配置された複数のストレージ装置の前記ファイルを管理する分散ストレージシステムのファイル管理方法において、

前記ストレージ装置が、新たにファイルを作成する度に前記管理サーバから割り当てられた一意の識別子としてのGUIDを当該作成するファイルに識別子として付加する処理と、

前記ストレージ装置が、当該ストレージ装置に割り当てられた一意の識別子としてのGUIDが不足した場合には、前記管理サーバにGUIDを要求する処理と、

前記管理サーバが、前記ストレージ装置からの前記要求に基づいて、未使用のGUIDをブロック単位で前記ストレージ装置に割り当て、当該ストレージ装置の装置IDを前記GUIDに割り当てる処理と、

前記管理サーバが、前記ストレージ装置に割り当てた前記GUIDと、前記ストレージ装置の装置IDの関係を当該管理サーバのGUID割当データベースに記憶する処理と、

前記管理サーバは、一つまたは複数のグループと、これら各グループに所属するストレージ装置の装置IDとグループの識別子の対応関係を管理するグループ管理データベースを有し、前記ストレージ装置を追加する際には、当該ストレージ装置に前記装置IDを設定し、前記装置IDに前記グループの識別子を関連付けて前記グループ管理データベース

10

20

に登録する処理と、

前記管理サーバは、前記ストレージ装置に割り当てた未使用の G U I D 及びストレージ装置の装置 I D を、前記割り当てを行ったストレージ装置が所属するグループのストレージ装置に送信する処理と、

前記グループに所属するストレージ装置が、前記 G U I D 及び装置 I D を受信し、それぞれに設けたローカルの G U I D 管理データベースに前記送信された G U I D 及びストレージ装置の装置 I D を G U I D 管理データベースに格納する処理と、

前記ストレージ装置は、異なるグループのストレージ装置からのアクセスを受信したときには当該アクセスに対してエラーを通知する処理と、

前記ストレージ装置は、同一のグループのストレージ装置からのアクセスを受信したときには前記 G U I D 管理データベースの G U I D と装置 I D の関係から、前記アクセスに含まれるファイルのオリジナルを格納する装置 I D を選択し、当該装置 I D のストレージ装置に対して前記アクセスを行う処理と、

前記ストレージ装置は、当該ストレージ装置の負荷を検出し、前記検出した負荷が予め設定した値以下のときにはストレージ装置間でレプリカを作成する処理と、

を含み、

前記ストレージ装置間でレプリカを作成する処理は、

前記ストレージ装置は、ローカルに設けた G U I D 管理データベースの G U I D と装置 I D の関係から、オリジナルのファイルを選択する処理と、

前記選択したオリジナルのファイルを、他のストレージ装置へ複製する処理と、

を含むことを特徴とする分散ストレージ装置のファイル管理方法。

【請求項 2】

前記管理サーバは、前記割り当てた未使用の G U I D 及びストレージ装置の装置 I D を、各ストレージ装置に送信する処理と、

各ストレージ装置は、それぞれに設けたローカルの G U I D 管理データベースに前記送信された G U I D 及びストレージ装置の装置 I D を格納する処理と、

を含むことを特徴とする請求項 1 に記載の分散ストレージ装置のファイル管理方法。

【請求項 3】

前記ストレージ装置は、同一のグループのストレージ装置からのアクセスを受信したときには前記 G U I D 管理データベースの G U I D と装置 I D の関係から、前記アクセスに含まれるファイルのオリジナルを格納する装置 I D を選択し、当該装置 I D のストレージ装置に対して前記アクセスを行う処理は、

前記ストレージ装置が、クライアント装置から当該ストレージ装置に対するファイルの書き込み要求を受信する処理と、

前記ストレージ装置が G U I D 管理データベースの G U I D と装置 I D の関係から、オリジナルのファイルを有するストレージ装置を選択する処理と、

前記ストレージ装置が前記選択したストレージ装置に対して書き込む処理と、

を含むことを特徴とする請求項 1 に記載の分散ストレージ装置のファイル管理方法。

【請求項 4】

前記ストレージ装置が前記選択したストレージ装置に対して書き込む処理は、

前記オリジナルのファイルを有するストレージ装置が、複数のクライアント装置から書き込み要求を受信した場合には、各書き込み要求を逐次処理し、各書き込み処理が終了するたびに前記クライアント装置へ応答を返すことを特徴とする請求項 3 に記載の分散ストレージ装置のファイル管理方法。

【請求項 5】

前記ストレージ装置は、同一のグループのストレージ装置からのアクセスを受信したときには前記 G U I D 管理データベースの G U I D と装置 I D の関係から、前記アクセスに含まれるファイルのオリジナルを格納する装置 I D を選択し、当該装置 I D のストレージ装置に対して前記アクセスを行う処理は、

前記ストレージ装置が、クライアント装置から当該ストレージ装置に対するファイルの

10

20

30

40

50

読み込み要求を受信する処理と、

当該ストレージ装置に前記要求されたファイルがない場合には、前記ファイルに対応する G U I D に基づいて当該ストレージ装置の G U I D 管理データベースからオリジナルのファイルを有するストレージ装置の装置 I D を選択し、この選択した装置 I D に対応するストレージ装置から前記要求されたファイルを読み込む処理と、
を含むことを特徴とする請求項 1 に記載の分散ストレージ装置のファイル管理方法。

【請求項 6】

前記ストレージ装置は、同一のグループのストレージ装置からのアクセスを受信したときには前記 G U I D 管理データベースの G U I D と装置 I D の関係から、前記アクセスに含まれるファイルのオリジナルを格納する装置 I D を選択し、当該装置 I D のストレージ装置に対して前記アクセスを行う処理は、

10

前記ストレージ装置が、クライアント装置から当該ストレージ装置に対するファイルの読み込み要求を受信する処理と、

当該ストレージ装置に前記要求されたファイルがない場合には、前記ファイルに対応する G U I D に基づいて当該ストレージ装置の G U I D 管理データベースからオリジナルのファイルを有するストレージ装置を選択する処理と、

前記選択したストレージ装置に対して前記読み込み要求を転送し、要求されたファイルを読み込む処理と、

この読み込んだファイルを前記クライアント装置に送る処理と、

を含むことを特徴とする請求項 1 に記載の分散ストレージ装置のファイル管理方法。

20

【請求項 7】

分散配置された複数のストレージ装置と、

前記ストレージ装置を管理する管理サーバと、を備え、

前記ストレージ装置が、前記ストレージ装置間でファイルの複製を作成するレプリケーション手段を有する分散ストレージシステムにおいて、

前記ストレージ装置は、

当該ストレージ装置で作成したファイルに分散ストレージシステム内で一意の識別子となる G U I D を付与する G U I D 付与手段と、

前記ストレージ装置に格納したファイルの G U I D とオリジナルのファイルを有するストレージ装置の装置 I D との関連を格納する G U I D 管理データベースと、

30

クライアント装置からのファイルのアクセス要求に応じてファイルを検索するファイル検索手段と、

前記クライアント装置から要求されたファイルが当該ストレージ装置に無い場合には、前記 G U I D 管理データベースから前記ファイルの G U I D に基づいてオリジナルのファイルを有するストレージ装置の装置 I D を検索する基準ストレージ装置検索手段と、

前記検索した装置 I D のストレージ装置に前記ファイルのアクセスを要求し、前記要求されたファイルにアクセスするファイルアクセス手段と、

前記ストレージ装置の負荷を検出し、前記検出した負荷が予め設定した値以下のときにはストレージ装置間でレプリカを作成するレプリケーション手段と、

を有し、

40

前記管理サーバは、

分散ストレージシステム内で一意の識別子となる G U I D とストレージ装置の装置 I D との関係を格納する G U I D 割当データベースを有し、前記ストレージ装置からの前記要求に基づいて、未使用の G U I D をブロック単位で前記ストレージ装置に割り当て、当該ストレージ装置の装置 I D を前記 G U I D に割り当て、

一つまたは複数のグループと、これら各グループに所属するストレージ装置の装置 I D とグループの対応関係を管理するグループ管理データベースを有し、前記ストレージ装置を追加する際には、当該ストレージ装置に前記装置 I D を設定し、前記装置 I D に前記グループを関連付けて前記グループ管理データベースに登録し、前記割り当てた未使用の G U I D 及びストレージ装置の装置 I D を、前記割り当てを行ったストレージ装置が所属す

50

るグループのストレージ装置に送信し、

前記 G U I D 付与手段は、

新たにファイルを作成する度に前記 G U I D 管理データベースから未使用の G U I D を当該作成するファイルに識別子として付加し、当該ストレージ装置に割り当てられた G U I D が不足した場合には、前記管理サーバに G U I D を要求し、前記管理サーバから割り当てられた G U I D と装置 I D を前記 G U I D 管理データベースに格納し、

前記ファイル検索手段は、

異なるグループのストレージ装置からのアクセスを受信したときには当該アクセスに対してエラーを通知し、

前記レプリケーション手段は、

前記ストレージ装置に設けた前記 G U I D 管理データベースの G U I D と装置 I D の関係から、オリジナルのファイルを選択し、前記選択したオリジナルのファイルを、他のストレージ装置へ複写することを特徴とする分散ストレージシステム。

【請求項 8】

分散配置された複数のストレージ装置と、

これらストレージ装置が格納するファイルの識別子及びストレージ装置を管理する管理サーバとを備えて、前記分散配置された複数のストレージ装置の前記ファイルを管理する分散ストレージシステムのファイル管理方法において、

前記ストレージ装置が、新たにファイルを作成する度に前記管理サーバから割り当てられた一意の識別子としての G U I D を当該作成するファイルに識別子として付加する処理と、

前記ストレージ装置が、当該ストレージ装置に割り当てられた一意の識別子としての G U I D が不足した場合には、前記管理サーバに G U I D を要求する処理と、

前記管理サーバが、前記ストレージ装置からの前記要求に基づいて、未使用の G U I D をブロック単位で前記ストレージ装置に割り当て、当該ストレージ装置の装置 I D を前記 G U I D に割り当てる処理と、

前記管理サーバが、前記ストレージ装置に割り当てた前記 G U I D と、前記ストレージ装置の装置 I D の関係を当該管理サーバの G U I D 割当データベースに記憶する処理と、

前記管理サーバは、一つまたは複数のグループと、これら各グループに所属するストレージ装置の装置 I D とグループの識別子の対応関係を管理するグループ管理データベースを有し、前記ストレージ装置を追加する際には、当該ストレージ装置に前記装置 I D を設定し、前記装置 I D に前記グループの識別子を関連付けて前記グループ管理データベースに登録する処理と、

前記管理サーバは、前記ストレージ装置に割り当てた未使用の G U I D 及びストレージ装置の装置 I D を、前記割り当てを行ったストレージ装置が所属するグループのストレージ装置に送信する処理と、

前記グループに所属するストレージ装置が、前記 G U I D 及び装置 I D を受信し、それぞれに設けたローカルの G U I D 管理データベースに前記送信された G U I D 及びストレージ装置の装置 I D を G U I D 管理データベースに格納する処理と、

前記ストレージ装置が、ファイルに対するアクセス要求を受信したときには、前記ファイルが当該ストレージ装置に格納されているか否かを判定する処理と、

前記ストレージ装置が、前記ファイルが当該ストレージ装置に格納されていると判定した場合には、当該ファイルに対するアクセスを行う処理と、

前記ストレージ装置が、前記ファイルが当該ストレージ装置に格納されていないと判定した場合には、当該ファイルの G U I D から G U I D 管理データベースを参照して当該ファイルを格納する装置 I D を検索する処理と、

前記ストレージ装置が、前記 G U I D に対応する装置 I D を取得した場合には、取得した装置 I D のストレージ装置に対して前記アクセスを行う処理と、

前記ストレージ装置が、前記 G U I D に対応する装置 I D を取得できない場合には、前記アクセスに対してエラーを通知する処理と、

10

20

30

40

50

前記ストレージ装置は、当該ストレージ装置の負荷を検出し、前記検出した負荷が予め設定した値以下のときにはストレージ装置間でレプリカを作成する処理と、

を含み、

前記ストレージ装置間でレプリカを作成する処理は、

前記ストレージ装置は、ローカルに設けたGUID管理データベースのGUIDと装置IDの関係から、オリジナルのファイルを選択する処理と、

前記選択したオリジナルのファイルを、他のストレージ装置へ複写する処理と、

を含むことを特徴とする分散ストレージ装置のファイル管理方法。

【発明の詳細な説明】

【技術分野】

10

【0001】

本発明は、ネットワークを介して記憶装置を分散して配置し、各記憶装置に格納されたファイルを共有する分散ストレージシステムの改良に関する。

【背景技術】

【0002】

クライアント/サーバ型の分散ストレージシステム（または装置）としては、Andrew File System（以下、AFSという）が知られている（非特許文献1）。

【0003】

これは、各ファイルにロケーションフィールドを備えた識別子を付加し、このファイルの位置をロケーションフィールドの情報により取得して、目的とするファイルにアクセスするものである。このため、ロケーションフィールドを管理するロケーションデータベースを任意のネットワーク上に設置しておき、クライアントはファイルにアクセスする際に、まず、識別子に含まれたロケーションフィールドについてロケーションデータベースへ問い合わせを行い、ロケーションデータベースが回答した位置に対してアクセスを行っている。

20

【0004】

そして、一回アクセスされたファイルは、ストレージ装置やキャッシュサーバなどにキャッシュされ、次回からのアクセス速度を向上させている。

【非特許文献1】「最前線UNIXのカーネル」 ユーレッシュ・ヴァハリア著、徳田英幸、中村明、戸部義人、津田悦幸 訳、株式会社ピアソン・エデュケーション、2000年5月15日発行、第371頁～378頁、（原著 UNIX Internals:The New Frontiers by Uresh Vahalia published by Prentice Hall, Inc. 1996）

30

【発明の開示】

【発明が解決しようとする課題】

【0005】

しかしながら、上記従来例では、ローカルのストレージ装置にファイルがある場合には、迅速にファイルの読み書きを行うことができるが、ローカルのストレージ装置にファイルがない場合には、ロケーションデータベースへ問い合わせを行い、ロケーションデータベースからの回答を待って、目的のストレージ装置にリモートアクセスすることになり、このリモートアクセスを迅速に行うことができないという問題があった。

40

【0006】

また、リモートアクセスによりキャッシュへアクセスする場合には、データの整合性を確保するために排他制御を必要としているため、可用性が低下するという問題がある。例えば同一のファイルに対して複数の書き込み要求があったときには、最先のクライアントの書き込み要求を受け付けて、後の書き込み要求を拒否することになる。

【0007】

さらに、基準となるストレージ装置のオリジナルのデータを、リモートにある他のストレージ装置に複写するレプリケーションを行う場合、レプリケーションの完了までには時間を要し、レプリケーションが完了する以前に複写先のストレージ装置にアクセスが発生すると、ファイルの複製（レプリカ）が完成していないため、アクセス要求がエラーとな

50

ってしまい、信頼性が低下するという問題があった。

【0008】

そこで本発明は、上記問題点に鑑みてなされたもので、レプリケーション中であっても整合性を保ちながらファイルのアクセスを可能にすることを目的とする。

【課題を解決するための手段】

【0009】

本発明は、分散配置された複数のストレージ装置と、これらストレージ装置が格納するファイルの識別子及びストレージ装置を管理する管理サーバとを備えて前記分散配置された複数のストレージ装置の前記ファイルを管理する分散ストレージシステムのファイル管理方法において、前記ストレージ装置が、新たにファイルを作成する度に前記管理サーバから割り当てられた一意の識別子としてのGUIDを当該作成するファイルに識別子として付加する処理と、前記ストレージ装置が、当該ストレージ装置に割り当てられた一意の識別子としてのGUIDが不足した場合には、前記管理サーバにGUIDを要求する処理と、前記管理サーバが、前記ストレージ装置からの前記要求に基づいて、未使用のGUIDをブロック単位で前記ストレージ装置に割り当て、当該ストレージ装置の装置IDを前記GUIDに割り当てての処理と、前記管理サーバが、前記ストレージ装置に割り当てた前記GUIDと、前記ストレージ装置の装置IDの関係を当該管理サーバのGUID割当データベースに記憶する処理と、前記管理サーバは、一つまたは複数のグループと、これら各グループに所属するストレージ装置の装置IDとグループの識別子の対応関係を管理するグループ管理データベースを有し、前記ストレージ装置を追加する際には、当該ストレージ装置に前記装置IDを設定し、前記装置IDに前記グループの識別子を関連付けて前記グループ管理データベースに登録する処理と、前記管理サーバは、前記ストレージ装置に割り当てた未使用のGUID及びストレージ装置の装置IDを、前記割り当てを行ったストレージ装置が所属するグループのストレージ装置に送信する処理と、前記グループに所属するストレージ装置が、前記GUID及び装置IDを受信し、それぞれに設けたローカルのGUID管理データベースに前記送信されたGUID及びストレージ装置の装置IDをGUID管理データベースに格納する処理と、前記ストレージ装置は、異なるグループのストレージ装置からのアクセスを受信したときには当該アクセスに対してエラーを通知する処理と、前記ストレージ装置は、同一のグループのストレージ装置からのアクセスを受信したときには前記GUID管理データベースのGUIDと装置IDの関係から、前記アクセスに含まれるファイルのオリジナルを格納する装置IDを選択し、当該装置IDのストレージ装置に対して前記アクセスを行う処理と、前記ストレージ装置は、当該ストレージ装置の負荷を検出し、前記検出した負荷が予め設定した値以下のときにはストレージ装置間でレプリカを作成する処理と、を含み、前記ストレージ装置間でレプリカを作成する処理は、前記ストレージ装置は、ローカルに設けたGUID管理データベースのGUIDと装置IDの関係から、オリジナルのファイルを選択する処理と、前記選択したオリジナルのファイルを、他のストレージ装置へ複写する処理と、を含む。

【発明の効果】

【0010】

したがって、本発明は、ストレージ装置間でレプリケーションが行われているとき、ストレージ装置に複製が作成される以前にクライアント装置からアクセスがあったときには、ファイルに付加される一意の識別子であるGUIDに基づいてオリジナルのファイルを有するストレージ装置を検索し、このストレージ装置のオリジナルのファイルにアクセスすることで、確実にクライアント装置からの要求に応えることができ、レプリケーションを行う分散ストレージシステムの信頼性を向上することができる。また、クライアント装置はオリジナルのファイルにアクセス可能となるので、ファイルの整合性を保証できる。

【発明を実施するための最良の形態】

【0011】

以下、本発明の一実施形態を添付図面に基づいて説明する。

【0012】

10

20

30

40

50

図1は、本発明を適用する分散ストレージシステムの構成図である。この図1では、2つのストレージ装置0、1とNISサーバ(Network Information Server)1がネットワーク10を介して接続され、分散ストレージシステムを構築した例を示している。なお、ネットワーク10はインターネット、WAN、LANなどを示す。

【0013】

ストレージ装置0は、データとしてのファイルを格納するディスク14と、ディスク14に格納されるファイルに付与するGUID(Globally Unique Identifier)の情報を管理するGUID管理データベース(以下、GUID管理DBという)13を備える。GUID管理DB13はディスク14に格納されても良いし、ストレージ装置0の他のディスク(図示省略)に格納されても良い。

10

【0014】

なお、ファイルに付加されるGUIDは、この分散ストレージシステム上でファイルを一意に識別するための識別子であり、後述するように、例えば、番号などで構成される。

【0015】

これら、GUID管理DB13、ディスク14は、CPU(図示省略)やメモリ、I/Oなどを含む制御部12によって、ネットワーク10やクライアント装置との間で読み書きの制御が行う。また、制御部12には、後述する書き込み要求の際に使用するバッファ16を備える。

【0016】

さらに、制御部12には、ディスク14または制御部12自身の処理負荷を測定する負荷検出装置15が設けられ、レプリケーションを実行する際には負荷検出装置15が検出した負荷を判定し、検出した負荷が所定値以下のときなどにレプリケーションを実行する。

20

【0017】

また、ストレージ装置0にはCPUやメモリ、コンソールなど(図示省略)を含むクライアント装置10が接続され、クライアント装置10からはストレージ装置0をローカルなストレージとして利用することができる。なお、図中ストレージ装置0がネットワーク10に接続されている例を示したが、クライアント装置10またはストレージ装置0の少なくとも一方がネットワーク10に接続されていればよい。

【0018】

ネットワーク10に接続されたストレージ装置1も、上記ストレージ装置0と同様に構成され、ファイルを格納するストレージ24、ファイルに付与するGUIDの情報を管理するGUID管理DB23、制御部22、バッファ26、負荷検出装置25から構成され、また、ストレージ装置1にはCPUやメモリ、コンソールなど(図示省略)を含むクライアント装置11が接続され、クライアント装置11からはストレージ装置1をローカルなストレージとして利用することができる。

30

【0019】

このストレージ装置1も、上記ストレージ装置0と同様に、ネットワーク10またはクライアント装置11からのアクセスに応じてストレージ24やGUID管理DB23との間で読み書きを行う。

40

【0020】

そして、ストレージ装置0とストレージ装置1は、クライアント装置10、11からのローカルまたはネットワーク10を介したりリモートのアクセスに応じてディスク14、24の読み書きを行う。そして、ストレージ装置0または1の負荷が低いときには、各ストレージ装置0、1が格納しているオリジナルのファイルを他方のストレージ装置に複写するレプリケーションを実行する。

【0021】

ネットワーク10には、GUIDとストレージ装置0、1を管理するNISサーバ1が接続される。なお、NISサーバは必ずしもサーバとして独立している必要はなく、ストレージ装置内またはネットワーク装置内にその機能を持たせても良い。

50

【 0 0 2 2 】

N I Sサーバ1は、C P Uやメモリ、コンソールなど（図示省略）からなる構成された制御部2と、G U I Dの割当状況を管理するG U I D割当データベース（以下、G U I D割当DBという）3から構成される。

【 0 0 2 3 】

N I Sサーバ1は、ネットワーク10上に接続され、かつ、分散ストレージシステムを構成するストレージ装置の登録、管理を行うとともに、各ストレージ装置0、1が格納するファイルに付与するG U I Dの割り当てを行うものである。

【 0 0 2 4 】

なお、図1においては、ストレージ装置0、1がシステム管理者などによって、N I Sサーバ1に登録された状態を示している。

10

【 0 0 2 5 】

次に、図2、図3を参照しながらG U I Dとファイル及びストレージ装置の関係を説明する。

【 0 0 2 6 】

図2は、ストレージ装置0のディスク14に、ファイルA、Bが既に格納されており、新たにファイルCを作成する状態を示している。また、ストレージ装置1のディスク24にファイルD、E、Fを既に格納した状態を示している。

【 0 0 2 7 】

ストレージ装置0、1は、格納するファイルを分散ストレージシステム上で一意に特定するため、N I Sサーバ1から割り当てられたG U I Dを各ファイルに付加している。例えば、ストレージ装置0のファイルAにはG U I D = 1 0 0 0 1が、ファイルBにはG U I D = 1 0 0 0 2が付加され、同様にストレージ装置1のファイルD ~ Fには、G U I D = 2 0 0 0 1 ~ 2 0 0 0 3が付加されている。

20

【 0 0 2 8 】

これらG U I Dは、N I Sサーバ1が予め設定したブロック単位で各ストレージ装置に割り当てるもので、各ストレージ装置は、ローカルのG U I D管理DB13、23に格納されたG U I Dのうち未使用のG U I Dが不足すると、N I Sサーバ1に対してG U I Dの割り当てを要求する。

【 0 0 2 9 】

要求を受けたN I Sサーバ1は、未使用のG U I Dをブロック単位でストレージ装置に付与し、N I Sサーバ1のG U I D割当DB3を更新した後、分散ストレージシステムを構成するストレージ装置0、1に対して、G U I Dブロックを割り当てたストレージ装置番号（または装置ID）をブロードキャストする。なお、G U I Dのブロックは、予め設定した範囲で連続した識別子の一群を意味し、例えば、図示のように、連続した1 0 0 0 0の番号をひとつのブロックとして扱う。

30

【 0 0 3 0 】

各ストレージ装置0、1は、N I Sサーバ1からブロードキャストされたG U I Dブロックとストレージ装置番号を、各G U I D管理DB13、23に格納する。これにより、各ストレージ装置は、アクセス要求があったときに、ファイルに対応するG U I Dから、どのストレージ装置にオリジナルのファイルがあるかを特定することができる。

40

【 0 0 3 1 】

今、図2において、ストレージ装置0が新たなファイルCを作成し、既に割り当てられたG U I Dブロック（1 0 0 0 1 ~ 2 0 0 0 0）を使用してしまった場合、ストレージ装置0はG U I Dのブロック割当をN I Sサーバ1に要求する（図中S1）。

【 0 0 3 2 】

要求を受けた、N I Sサーバ1は、G U I D割当DB3を調べ、一つ目のブロック1 0 0 0 1 ~ 2 0 0 0 0がストレージ装置0に、二つ目のブロック2 0 0 0 1 ~ 3 0 0 0 0がストレージ装置1に割り当てられているので、未使用である三つ目のG U I Dブロック3 0 0 0 1 ~ 4 0 0 0 0をストレージ装置0に割り当て、通知する（図中S2）。

50

【 0 0 3 3 】

通知を受けたストレージ装置 0 は、新たなファイル C に対して未使用且つ最小の G U I D = 3 0 0 0 1 を付加する。

【 0 0 3 4 】

なお、未使用 G U I D が枯渇する前に少なくなった時点で予め新しい G U I D ブロックの割当てを要求してもよい。またファイルに付加する G U I D は未使用なものであれば、必ずしも最小でなくとも良い。

【 0 0 3 5 】

次に、N I S サーバ 1 は、図 3 に示すように、G U I D 割当 D B 3 に対して、新たに割り当てた G U I D ブロック 3 0 0 0 1 ~ 4 0 0 0 0 に対応するストレージ装置番号に 0 を登録する (図中 S 3) 。

10

【 0 0 3 6 】

次に、N I S サーバ 1 は、分散ストレージシステムに所属する全てのストレージ装置 (ここでは 0、 1) に対して、更新した G U I D 割当 D B 3 の情報 (割り当てた G U I D ブロックと要求したストレージ装置番号の組み) をブロードキャストする (図中 S 4) 。

【 0 0 3 7 】

N I S サーバ 1 から G U I D 割当 D B の更新情報を受信した各ストレージ装置 0、 1 は、ローカルの G U I D 管理 D B 1 3、 2 3 を更新し、G U I D ブロック 3 0 0 0 1 ~ 4 0 0 0 0 がストレージ装置 0 に割り当てられたことを登録する (図中 S 5) 。

20

【 0 0 3 8 】

この G U I D ブロックの割り当てを時系列で見ると、図 4 のようになり、ストレージ装置 0 が G U I D ブロックを要求すると、N I S サーバ 1 は S 2 で通知を行うため、図中 S 6 でストレージ装置 0 は新たなファイル C に G U I D を付加することができる。

【 0 0 3 9 】

その後、N I S サーバ 1 は全てのストレージ装置に対して更新した情報をブロードキャストするので、図中 S 5 以降では、全てのストレージ装置の G U I D 管理 D B 1 3、 2 3 は同一の内容となる。このため、クライアント装置 1 0 または 1 1 がストレージ装置に対してアクセス要求をすると、どのストレージ装置に対するアクセス要求も同一のストレージ装置の内容を伝えるため、データの整合性を確保できるのである。

30

【 0 0 4 0 】

以上のように、G U I D 管理 D B 1 3、 2 3 と G U I D 割当 D B 3 の G U I D ブロックとストレージ装置番号 (装置 I D) により、ストレージ装置 0、 1 内のファイルは、G U I D によりオリジナルのファイルを格納しているストレージ装置番号を一意に特定することができ、どのストレージ装置からもファイルのロケーションを迅速に問い合わせることが可能となるのである。

【 0 0 4 1 】

そして、ファイルを作成したときに付与する G U I D は、管理サーバ 1 から各ストレージ装置 0、 1 に対してブロードキャストされ、各ストレージ装置の G U I D 管理 D B 1 3、 2 3 に記録されるため、分散ストレージシステム内にはオリジナルファイルは一つであり、各 G U I D 管理 D B 1 3、 2 3 は、G U I D に基づいてオリジナルのファイルを有するストレージ装置の装置 I D を示すものである。

40

【 0 0 4 2 】

なお、G U I D ブロックを要求したストレージ装置に対しては、割り当ての通知 (S 2) とブロードキャスト (S 4) で、同一の内容の情報が 2 回 (S 2、 S 4) 通知されることになるが、G U I D 管理 D B の更新はどちらか一方の通知に基づいて行えばよい。

【 0 0 4 3 】

また、N I S サーバ 1 及びストレージ装置 0、 1 の G U I D 割当 D B 3 及び G U I D 管理 D B 1 3、 2 3 は、ストレージ装置番号と G U I D ブロックを対応付けて管理する例を示したが、図 5 で示すように、ストレージ装置番号とネットワークアドレスなどの位

50

置情報を示すストレージ識別子を併用しても良く、あるいは、GUIDブロックとストレージ識別子で管理を行っても良い。

【0044】

次に、分散ストレージシステムのレプリケーションについて図6を参照しながら説明する。

【0045】

図6は、ストレージ装置0のオリジナルのファイルA、B、Cのレプリカ(複製)をストレージ装置1に作成する一例を示す。

【0046】

ストレージ装置0は、負荷検出装置15が検出した負荷が予め設定した値以下の低負荷時(例えば、深夜などアクセスの少ない状態)であれば、ディスク14内のファイルのうちオリジナルのファイルA~Cについて他のストレージ装置1にレプリカを作成するレプリケーションを実行し、耐障害性を向上させる。なお、レプリケーションの実行は、オリジナルファイルを持つ基準ストレージ装置と、レプリカを格納するストレージ装置1の負荷が共に低いことが望ましい。なお、負荷検出装置15が検出する負荷は、例えば、制御部12のCPUの使用率、ディスク14の稼働量(単位時間当たりのRead/Writeバイト数)等を検出すればよい。

10

【0047】

ここで、基準ストレージ装置とは、オリジナルのファイルを格納しているストレージ装置を示し、GUID管理DB13、23またはGUID割当DB3のGUIDブロックとストレージ装置番号の組みが基準ストレージ装置を示している。例えば、GUIDが10001~20000のファイルであれば、ストレージ装置0が基準ストレージ装置であり、GUIDが20001~30000のファイルであれば、ストレージ装置1が基準ストレージ装置となる。

20

【0048】

レプリケーションは、GUID管理DB13を参照して、ストレージ装置番号が自らの番号と一致するGUIDを持つファイルを、他のストレージ装置に複製することで行われる。

【0049】

したがって、ストレージ装置0のレプリケーションは、GUIDブロック10001~20000の範囲にあるファイルA、Bと、GUIDブロック30001~40000の範囲にあるファイルCを、ストレージ装置1に複製することで行われる。

30

【0050】

ストレージ装置1のディスク24には、ファイルA~Cの複製A'~C'が作成される。ストレージ装置1に複製されたファイルA'~C'は説明のために「'」を付しただけであり、実際はファイルA~Cと同一であって、GUIDもストレージ装置0のファイルA~Cと同一である。

【0051】

次に、レプリケーションは開始から完了まで時間を要するため、図7のように、ストレージ装置0からストレージ装置1へレプリケーションを行っている期間で、図中波線で示すファイルCのレプリカがまだ完了していないとき、クライアント装置11がローカルのストレージ装置1に対してGUID=30001のファイルCに対して読み込み(READ)要求を行う場合がある。

40

【0052】

この場合、ストレージ装置1は次のような制御を行う。

【0053】

図7において、クライアント装置11からのファイルCへのアクセス要求から、ストレージ装置1は、ファイルC(図中波線)のレプリカがローカルのディスク24にまだ作成されていないことが分かる。ストレージ装置1は、GUID管理DB23からGUID30001のファイルCの基準ストレージ装置が何れであることを検索し、ストレージ

50

装置 0であることを判定する。

【0054】

ストレージ装置 1は、基準ストレージ装置であるストレージ装置 0に対してGUID = 30001のファイルCの読み込みを要求する。ストレージ装置 1は、ストレージ装置 0からオリジナルのファイルCを転送し、クライアント装置 11に送信する。

【0055】

こうして、分散ストレージシステムのレプリケーションでは、ネットワーク10等を介して位置的に異なるストレージ装置へレプリカを作成する際、全てのレプリカが完成するまでには遅延が発生するが、まだレプリカが作成されていないファイルにアクセス要求があったときには、オリジナルのファイルを格納した基準ストレージ装置へアクセス要求を転送し、基準ストレージ装置から読み出すことにより、レプリカの完成、未完成に係わらず、分散ストレージシステム上では必ず目的のファイルを読み出すことになって、データの整合性を確保できる。特に、頻繁にレプリケーションを行う場合であっても、クライアント装置 10やクライアント装置 11からアクセス要求があったとき、要求されたファイルが作成されていない場合は、GUIDにより一意で決まる基準ストレージ装置に対してアクセスが行われるので、レプリケーションによる遅延の影響を受けることなくファイルの読み出しを保証することができるのである。

【0056】

上記図7で示した読み出し要求に対するストレージ装置 1の制御の一例を図8に示す。この図8は、クライアント装置 10または 11からのコマンドを受けたときに、ストレージ装置 1の制御部22で実行されるフローチャートを示す。

【0057】

まず、S11では要求されたコマンドがREAD（読み出し）要求コマンドであるかを判定し、READコマンドであれば、S12へ進んでREADコマンドで要求されたファイルがディスク24にあるか否かを判定する。

【0058】

要求されたファイルがディスク24にある場合には、S20に進んで、ディスク24のファイルを要求先のクライアント装置に送信する。図7の例では、ファイルC以外の読み出し要求の場合、ファイルA'、B'、D～Fが読み出される。

【0059】

一方、上記S12の判定で、ディスク24には要求されたファイルがない場合、S13に進んで、要求されたファイルのGUIDに基づいてGUID管理DB23を検索し、基準ストレージ装置を読み込む。

【0060】

S14では、要求されたファイルの基準ストレージ装置の有無を判定し、基準ストレージ装置が見つからない場合には、S21に進んで要求先のクライアント装置に対してエラーを通知する。一方、基準ストレージ装置が見つかった場合には、S15に進んで、クライアント装置からのREADコマンドを基準ストレージ装置（図7の例ではストレージ装置 0）へファイルCの読み出し要求を転送する。

【0061】

S16では、ストレージ装置 1は、基準ストレージ装置であるストレージ装置 0から要求されたファイルCをリモートで読み出し、要求先のクライアント装置（クライアント装置 11）に送信する。

【0062】

以上の読み出し制御により、ストレージ装置 1は、クライアント装置からファイルの読み出し要求があると、ローカルのディスク24にファイルがあれば、このディスク24のファイルをクライアント装置に送信するが、ディスク24にファイルが無い、または作成途中の場合では、要求されたファイルのGUIDから、そのファイルの基準ストレージ装置を検索し、読み出し要求を基準ストレージ装置に送り、リモートアクセスにより要求されたファイルを読み出してクライアント装置へ送信する。

10

20

30

40

50

【 0 0 6 3 】

これにより、クライアント装置は、レプリケーション中の遅延の影響を受けることがなくなっており、迅速にファイルを読むことができ、レプリケーションを行う分散ストレージシステムのデータのアクセスを保証することができ、システムの信頼性を向上させるのである。

【 0 0 6 4 】

さらに、クライアント装置 10、11は、ストレージ装置 0、1に対して読み出し要求を行うだけであり、オリジナルのファイルがどこにあるかを認識する必要がなく、常に任意のストレージ装置に対してアクセスするだけで、分散ストレージシステム上のデータを取得することができ、クライアント装置からのアクセス性を向上することができる。

10

【 0 0 6 5 】

なお、上記ではストレージ装置 1に対してクライアント装置 11が読み出し要求を行った場合について説明したが、ストレージ装置 0に対してクライアント装置 10や11が読み出し要求を行った場合も上記と同様である。

【 0 0 6 6 】

また、上記ではクライアント装置 11からの読み出し要求を、ストレージ装置 1がオリジナルのファイルを持つストレージ装置 0に転送し、ストレージ装置 1がこれを読み込んでクライアント装置 11に送信する例を示したが、ストレージ装置 1がクライアント装置 11に基準ストレージ装置の位置を教え、クライアント装置 11が基準ストレージ装置にアクセスするようにしても良い。

20

【 0 0 6 7 】

次に、クライアント装置 10、11が同時に同一のファイルに対してWRITE（書き込み）要求を行った場合について、図9、図10を参照しながら説明する。

【 0 0 6 8 】

クライアント装置 10とクライアント装置 11は、同時にファイルAに対する書き込み要求を発生し、クライアント装置 10はローカルのストレージ装置 0に、クライアント装置 11はローカルのストレージ装置 1にそれぞれ要求する。

【 0 0 6 9 】

まず、ストレージ装置 0では、クライアント装置 10からの書き込み要求（S101）からファイルAの基準ストレージ装置をGUIDに基づいてGUID管理DB13から検索する。ファイルAの基準ストレージ装置はストレージ装置 0であるため、ストレージ装置 0はクライアント装置 10の要求を受け付け、制御部12のバッファ16にクライアント装置 10が送った新たなファイルAの内容を書き込む（S102）。バッファ16への書き込みが終了すると、ストレージ装置 0はクライアント装置 10に書き込み完了の通知を送信する（S103）。

30

【 0 0 7 0 】

一方、クライアント装置 11がストレージ装置 1に要求したファイルAの書き込みは、ストレージ装置 0に転送される。ストレージ装置 1には、ファイルAのレプリカA'が格納されているだけなので、レプリカに対する書き込みを禁止する。このため、ストレージ装置 1は、要求されたファイルAのGUIDに基づいて、GUID管理DB23からファイルAのオリジナルを持つ基準ストレージ装置を検索する。検索の結果、基準ストレージ装置がストレージ装置 0であることが解ると、ストレージ装置 1はストレージ装置 0に対してリモートアクセスによりファイルAの書き込みを要求する（S202）。

40

【 0 0 7 1 】

ストレージ装置 0は、クライアント装置 10からのバッファ16への書き込みが終了すると、ストレージ装置 1からの書き込み要求を受け付け、ストレージ装置 1はクライアント装置 11が要求したファイルAの内容をストレージ装置 0のバッファ16に書き込む（S203）。

50

【 0 0 7 2 】

ストレージ装置 0 は、クライアント装置 11 からのリモートアクセスによる書き込みが終了すると、ストレージ装置 1 に対して、書き込み完了の応答を送信する (S 2 0 4)。ストレージ装置 1 はこの応答を受信すると、クライアント装置 11 に対して書き込み完了の通知を送信する (S 2 0 5)。すなわち、クライアント装置に対しては、バッファ 16 への書き込みが完了した時点で完了の通知が送信される。

【 0 0 7 3 】

次に、ストレージ装置 0 は、バッファ 16 に書き込まれたファイル A の内容で、逐次ディスク 14 に書き込みを行う。つまり、最初に書き込み要求を受け付けたクライアント装置 10 の内容をファイル A に書き込んで更新し (S 1 0 4)、次に要求を受け付けたクライアント装置 11 の内容をファイル A に書き込んで更新する (S 2 0 6)。

10

【 0 0 7 4 】

こうして、書き込み要求に関しては、必ずオリジナルのファイルのファイルに対して処理を行うことで、分散して格納されたレプリカの整合性がとれなく無くなるのを防ぐことができ、ファイルの整合性を保証できる。

【 0 0 7 5 】

さらに、書き込み要求の処理は、要求を受け付けた順で、書き込むファイル A の内容をバッファ 16 に蓄積し、その後、バッファ 16 の内容でディスク 14 に書き込むようにしたため、書き込み順序の整合性を確保することができる。特に、ネットワーク 10 がインターネットで構成される場合は、データ (パケット) の到着順が要求の順序と一致しない場合も生じる。このため、一旦、ストレージ装置 0 の制御部 12 に設けたバッファ 16 に要求の受け付け順で、書き込む内容を蓄積し、バッファ 16 の書き込みが完了してからディスク 14 への書き込みを逐次行うので、ファイルの書き込み順序の整合性を保証できるのである。

20

【 0 0 7 6 】

なお、リモートアクセスによる書き込みとなるクライアント装置 11 では、完了の通知を受信するまでは、次の書き込み要求の送信を禁止するのが望ましい。つまり、上述のように、ネットワーク 10 がインターネットで構成される場合では、後の要求のパケットが、先の要求のパケットよりも先にストレージ装置 0 に到着する場合があるので、書き込み完了まで待ってから次の書き込み要求を送信することで、書き込み順序の不整合を防止できるのである。

30

【 0 0 7 7 】

図 11 は第 2 の実施形態を示し、前記第 1 実施形態のストレージ装置が、一つの分散ストレージシステムの中で一つのグループを構成していたのに対し、この第 2 実施形態では、一つの分散ストレージシステムの中に複数のグループを設け、各グループの中でアクセス及びレプリケーションを行うようにしたものである。

【 0 0 7 8 】

図 11 は、前記第 1 実施形態の図 2 に、ディスク 34 よ G U I D 管理 D B 33 を備えたストレージ装置 2 を加えるとともに、 N I S サーバ 1 には G U I D 割当 D B 3 に加えてグループ管理データベース (以下グループ管理 D B) 30 を加えたもので、その他の構成は前記第 1 実施形態に準ずる。

40

【 0 0 7 9 】

この分散ストレージシステムにおいては、 N I S サーバ 1 が管理するストレージ装置を 2 つのグループ A と B に分け、各グループの中で相互にアクセスするものであり、他のグループへのアクセスは禁止される。

【 0 0 8 0 】

このため、 N I S サーバ 1 には、グループ A とグループ B に所属するストレージ装置を管理するためのグループ管理 D B 30 を備える。グループ管理 D B 30 は、システム管理者等によって予め各グループに所属するストレージ装置番号 (またはストレージ識別子) が設定されており、例えば、グループ A にはストレージ装置 0 と 1 が属し、グループ

50

Bにはストレージ装置 2が所属するように設定されている。

【0081】

一方、GUID割当DB3が各ストレージ装置に割り当てるGUIDは、分散ストレージシステム全体で一意的識別子となるよう、グループに関係なく、ストレージ装置毎にGUIDブロックが割り当てられている。ストレージ装置 0、ストレージ装置 2には前記第1実施形態と同様に、10001～40000のGUIDブロックが割り当てられ、ストレージ装置 2には、40001～60000のGUIDブロックが割り当てられている。グループに拘わらず、GUIDをシステム全体で一意的識別子とすることで、ストレージ装置のグループを変更した場合に、GUIDが重複するのを防止することができる。

10

【0082】

今、ストレージ装置 0がGUIDブロックの割り当てをNISサーバ1に要求すると(S301)、NISサーバ1は前記第1実施形態と同様に、未使用のGUIDブロックをストレージ装置 0に割り当てる。

【0083】

その後、NISサーバ1は、新たなGUIDブロックに対応するストレージ装置番号をGUID割当DB3に登録して更新し、この更新の内容をストレージ装置 0が所属するグループAのみにブロードキャストする(S303)。

【0084】

つまり、NISサーバ1はグループ管理DB30を参照し、ストレージ装置 0が所属するグループAのストレージ装置 0、1のみに更新内容を通知し、グループAの各ストレージ装置 0、1は、ブロードキャストされた更新内容を各GUID管理DB13、23に登録する。

20

【0085】

これは、異なるグループ間でのアクセスは行わないため、グループBのストレージ装置は、グループAの基準ストレージ装置を指し示すGUIDを知る必要がないためである。

【0086】

次に、アクセス制御について、図12を参照しながら説明する。

【0087】

まず、クライアント装置 10がストレージ装置 1のファイルFを読み込む場合は、まず、ローカルに接続されたストレージ装置 0にファイルFの読み出し(READ)要求を送信する。ストレージ装置 0はディスク14にファイルFが無い場合、読み出し要求に含まれるファイルFのGUIDからGUID管理DB13を検索し、ファイルFの基準ストレージ装置であるストレージ装置 1を見つける。

30

【0088】

次に、ストレージ装置 0は、ストレージ装置 1に対してファイルFの読み出し要求を転送する(S312)。ストレージ装置 1はファイルFを読み出してストレージ装置 0にデータを転送する(S313)。最後にストレージ装置 0は、ストレージ装置 1から受信したファイルFのデータをクライアント装置 10に送信して処理を終了する。

40

【0089】

こうして、同一グループ内では、前記第1実施形態と同様にして、ストレージ装置のローカルのディスクにデータが無い場合には、GUIDに基づいて検索した基準ストレージ装置に問い合わせ、目的のファイルを読み出すのである。

【0090】

一方、異なるグループ間では、次のようになる。

【0091】

図12において、クライアント装置 11がローカルのストレージ装置 2に対してグループAのファイルF(GUID=20003)の読み出し要求を行った場合(S321)、ストレージ装置 2はローカルのディスク34にファイルFがないので、ファイルF

50

のGUIDに基づいてGUID管理DB33を検索する。

【0092】

しかし、GUID管理DB33には、グループAに関する情報はないため、基準ストレージ装置が見つからない。この結果、ストレージ装置2はクライアント装置11に対してエラーが生じたことを通知する(S322)。

【0093】

こうして、異なるグループ間では、各ストレージ装置のGUID管理DBに情報が登録されていないため、どのストレージ装置にアクセスすべきかを決定できず、相互のアクセスは禁止されるのである。

【0094】

これにより、同一の分散ストレージシステム内に、複数のグループを設けることで各グループ間のセキュリティを確保しながら、レプリケーションによって耐障害性を高めることが可能となる。

【0095】

なお、NISサーバ1のグループ管理DB30は、グループ名(グループ識別子)とストレージ装置番号とを対応付けて管理する例を示したが、図13で示すように、グループ名とストレージ装置のネットワークアドレスなどの位置情報を示すストレージIDによって管理しても良い。また、一つのストレージ装置を2つのグループに登録してもよく、例えば、ストレージ装置1をグループAとグループBに登録することで、2つのグループで一つのストレージ装置1を共有することもできる。

【0096】

例えば、図14で示すように、NISサーバ1のコンソールで、ストレージ装置の識別子をIPアドレスとし、複数のグループに登録する場合は、図中S401の画面にてストレージ装置のIPアドレスを入力した後、S402で登録するグループ名(グループ識別子)に「A」を入力し、S403では次に登録するグループ名として「B」を入力する。この結果、NISサーバ1のグループ管理DB30には、グループAとBにそれぞれこのストレージ装置を加えることができ、2つのグループでストレージ装置を共有することができる。

【0097】

なお、レプリケーションは、各グループ内でそれぞれ実行され、前記第1実施形態と同様に負荷が低いときに行えばよい。グループAでは、ストレージ装置0とストレージ装置1との間で、前記第1実施形態と同様に負荷が低いときに各オリジナルのファイルを他方へ複写する。図11、図12では説明を簡易にするためグループBのストレージ装置を2のみとしたが、グループBに複数のストレージ装置を設けて各ストレージ装置間でオリジナルのファイルを複写すればよい。

【0098】

なお、上記実施形態では、ストレージ装置0、1とした例を示したが、各ストレージ装置を複数のストレージ装置で構成されたストレージノードやサイトに置き換えても良い。

【産業上の利用可能性】

【0099】

以上のように、本発明に係る分散ストレージシステムでは、ローカルのストレージ装置にレプリケーションが完了していなくとも、ローカルのストレージ装置がオリジナルのファイルを有するストレージ装置からファイルを転送し、クライアント装置に読み出し可能とすることができるので、各ストレージ装置間のファイルの整合性を保証して可用性の高いサイトに適用することができる。

【図面の簡単な説明】

【0100】

【図1】本発明の一実施形態を示すシステム構成図。

【図2】GUIDの割り当ての様子を示す説明図。

10

20

30

40

50

【図3】GUIDの更新をブロードキャストする様子を示す説明図。

【図4】同じくGUIDの割り当てと更新のブロードキャストの様子を示すタイムチャート。

【図5】GUID割当データベースのデータの一例を示す説明図。

【図6】レプリケーションの様子を示す説明図。

【図7】レプリケーション中にクライアント装置から読み出し要求があった場合のデータの流れを示す説明図。

【図8】同じくストレージ装置で行われる読み出し処理の一例を示すフローチャート。

【図9】複数のクライアント装置から同時に書き込み要求があったときのデータの流れを示すタイムチャート。

【図10】同じく、複数のクライアント装置から同時に書き込み要求があったときのデータの流れを示す説明図。

【図11】第2の実施形態を示し、グループ単位でGUIDの割り当てと更新のブロードキャストを行う様子を示す説明図。

【図12】グループ単位で読み出し処理を行う際の、データの流れを示す説明図。

【図13】グループ管理データベースのデータの一例を示す説明図。

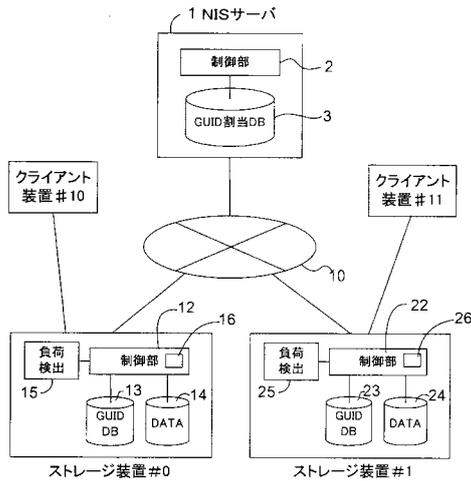
【図14】ストレージ装置を複数のグループで共有する場合の登録の様子を示す説明図。

【符号の説明】

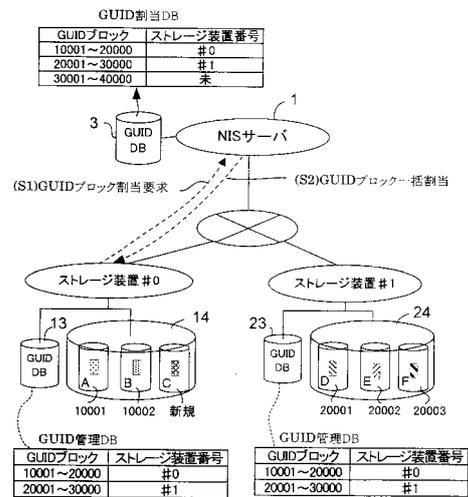
【0101】

- 1 NISサーバ
- 3 GUID割当データベース
- 13、23 GUID管理データベース
- 30 グループ管理データベース
- 0、1、2 ストレージ装置
- 10、11 クライアント装置

【図1】



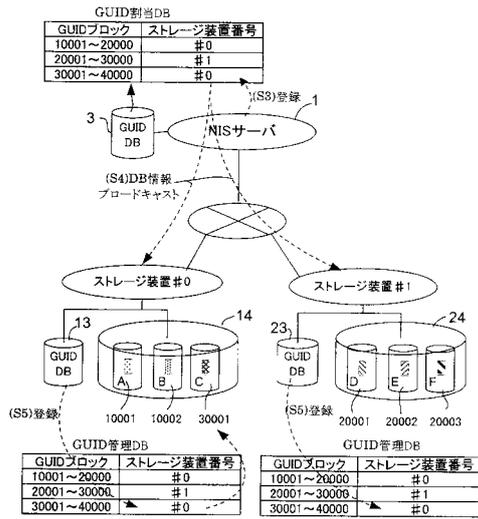
【図2】



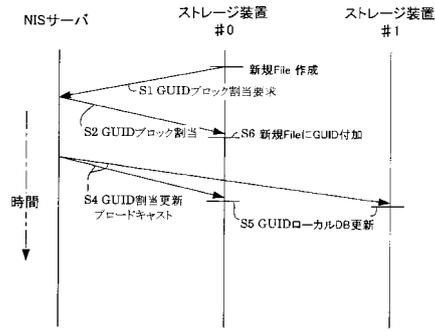
10

20

【図3】



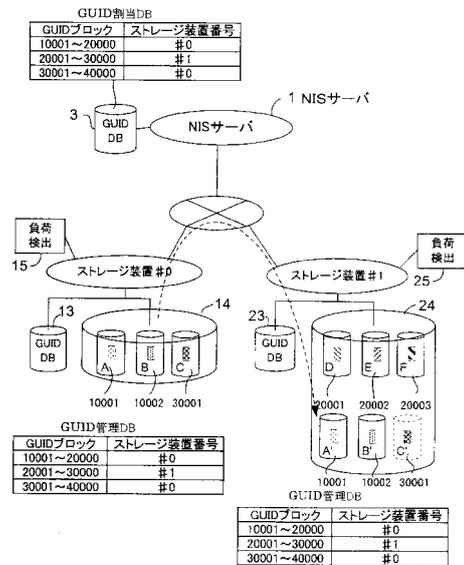
【図4】



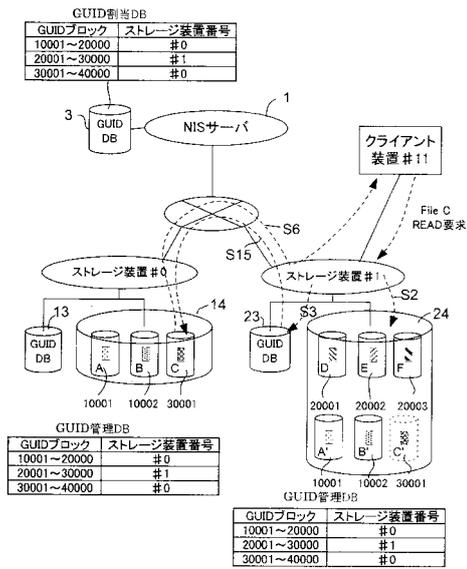
【図5】

GUIDブロック	ストレージ装置番号	ストレージ識別子
10001~20000	#0	172.29.1.1
20001~30000	#1	172.29.1.2
30001~40000	#0	172.29.1.1

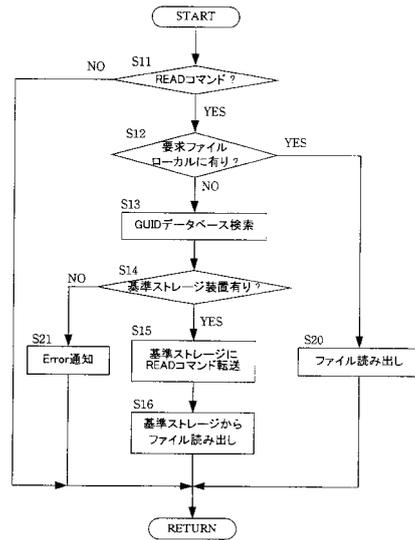
【図6】



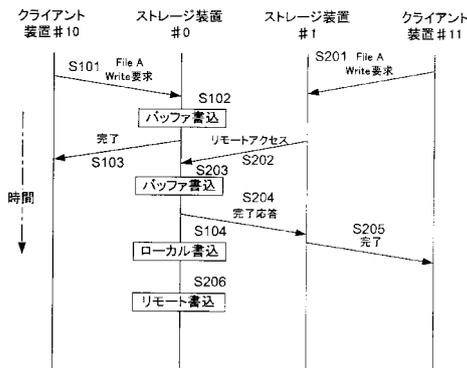
【図7】



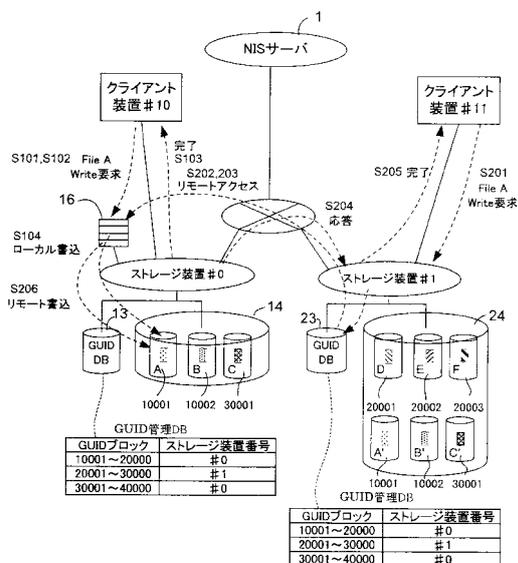
【図8】



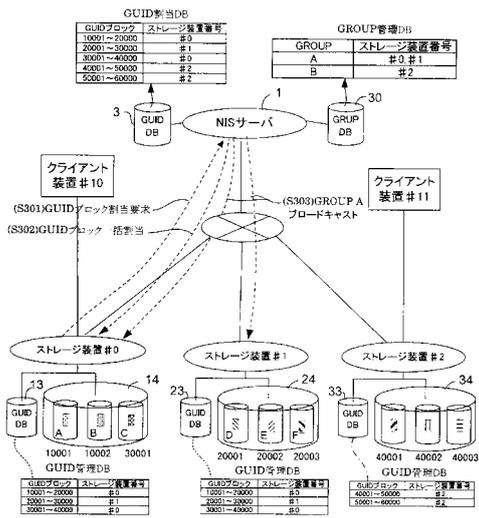
【図9】



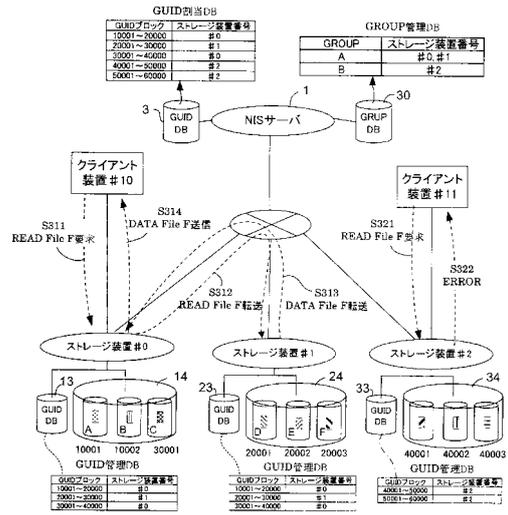
【図10】



【図 1 1】



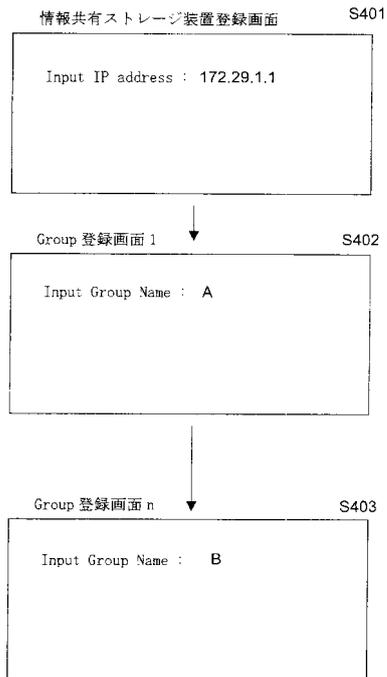
【図 1 2】



【図 1 3】

GROUP	ストレージ識別子
A	172.29.1.1, 172.29.1.2
B	172.29.2.1
C	172.29.2.2

【図 1 4】



フロントページの続き

(72)発明者 岩寄 正明

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

審査官 田川 泰宏

(56)参考文献 特開2003-173279(JP,A)

特開2002-264571(JP,A)

特開2002-245072(JP,A)

特開2003-131924(JP,A)

中内 清秀, ユビキタス環境に向けた分散コンテンツ発見, 電子情報通信学会技術研究報告, 日本, 社団法人電子情報通信学会, 2002年 9月23日, 第102巻, 第352号, p.7-12

(58)調査した分野(Int.Cl., DB名)

G06F 12/00