(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4699516号
(P4699516)
(45)発行日　平成23年6月15日(2011.6.15)　(24)登録日　平成23年3月11日(2011.3.11)

(51)Int.Cl.　F I
G06F 12/00 (2006.01)　G06F 12/00 501B
G06F 12/00 520E

請求項の数 10　(全 29 頁)

(21)出願番号　特願2008-507316(P2008-507316)
(86)(22)出願日　平成18年3月28日(2006.3.28)
(86)国際出願番号　PCT/JP2006/306284
(87)国際公開番号　WO2007/110931
(87)国際公開日　平成19年10月4日(2007.10.4)
審査請求日　平成20年6月25日(2008.6.25)

(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　100101856
弁理士　赤澤　日出夫
(72)発明者　塩沢　賢輔
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　新開　慶武
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　桜井　茂行

最終頁に続く

(54)【発明の名称】名前空間複製プログラム、名前空間複製装置、名前空間複製方法

(57)【特許請求の範囲】
【請求項１】
ストレージ装置上の名前空間の複製をコンピュータに実行させる名前空間複製プログラムであって、
前記ストレージ装置の制御を行うファイルシステム制御装置から前記名前空間の更新に関する情報である名前空間更新情報を取得し、前記ストレージ装置におけるファイル識別情報とリンク情報に基づいて作成されたデータベースである名前空間複製データベースを、前記名前空間更新情報に基づいて更新する名前空間複製データベース更新ステップと、
前記名前空間複製データベース更新ステップによる前記名前空間複製データベースへの反映が行われていない前記名前空間更新情報が喪失された場合、ファイル識別情報の更新時刻が所定時刻以降のファイル識別情報である未更新ファイル識別情報と、該未更新ファイル識別情報に対応するリンク情報である未更新リンク情報とを、前記ファイルシステム制御装置から取得し、前記未更新ファイル識別情報と前記未更新リンク情報とに基づいて前記名前空間複製データベースを修正する名前空間複製データベース修正ステップと
をコンピュータに実行させる名前空間複製プログラム。
【請求項２】
請求項１に記載の名前空間複製プログラムにおいて、
前記名前空間更新情報は、名前空間の更新の内容である名前空間更新内容と該更新の時刻である名前空間更新時刻を含み、
前記名前空間複製データベース修正ステップは、前記名前空間複製データベース更新ス

テップにより前記名前空間複製データベースに反映された名前空間更新情報に含まれる名前空間更新時刻のうち、最終のものを前記所定時刻とすることを特徴とする名前空間複製プログラム。

【請求項３】

請求項１に記載の名前空間複製プログラムにおいて、

前記名前空間複製データベース修正ステップは、前記未更新ファイル識別情報が示すファイルのうちディレクトリファイルの持つリンク情報を、前記未更新リンク情報とすることを特徴とする名前空間複製プログラム。

【請求項４】

請求項１に記載の名前空間複製プログラムにおいて、

前記名前空間複製データベース修正ステップは、前記所定時刻を前記ファイルシステム制御装置へ通知することにより、前記未更新ファイル識別情報と前記未更新リンク情報を抽出させ、取得することを特徴とする名前空間複製プログラム。

【請求項５】

請求項１に記載の名前空間複製プログラムにおいて、

前記名前空間複製データベース修正ステップは、前記名前空間複製データベースの修正が完了する前に、前記名前空間複製データベース更新ステップにより前記名前空間複製データベースに反映されていない名前空間更新情報を取得し、該名前空間更新情報が前記未更新ファイル識別情報に無関係である場合、取得した該名前空間更新情報に基づいて前記名前空間複製データベースを更新することを特徴とする名前空間複製プログラム。

【請求項６】

請求項１に記載の名前空間複製プログラムにおいて、

１つの前記リンク情報は、１つのディレクトリファイルのｉｎｏｄｅ情報と該ディレクトリファイルに含まれる子のファイルのｉｎｏｄｅ情報と該ディレクトリファイルに含まれる子のファイルの名前情報とを含み、

前記名前空間複製データベースは、前記リンク情報毎のエントリを持つことを特徴とする名前空間複製プログラム。

【請求項７】

請求項１に記載の名前空間複製プログラムにおいて、

前記ファイル識別情報は、ｉｎｏｄｅ情報であり、

前記リンク情報は、１つのディレクトリファイルのｉｎｏｄｅ番号と該ディレクトリファイルの子である１つのファイルのｉｎｏｄｅ番号とを含むことを特徴とする名前空間複製プログラム。

【請求項８】

請求項２に記載の名前空間複製プログラムにおいて、

前記名前空間複製データベース修正ステップは、前記名前空間複製データベースの修正が完了する前に、前記名前空間複製データベース更新ステップにより前記名前空間複製データベースに反映されていない名前空間更新情報を取得し、該名前空間更新情報が前記未更新ファイル識別情報に関係する場合、該名前空間更新情報の名前空間変更時刻と該名前空間更新情報に関係する前記未更新ファイル識別情報の更新時刻とを比較することにより、該名前空間更新情報と前記未更新ファイル識別情報のうち新しい方に基づいて前記名前空間複製データベースを修正することを特徴とする名前空間複製プログラム。

【請求項９】

ストレージ装置上の名前空間の複製を行う名前空間複製装置であって、

前記ストレージ装置の制御を行うファイルシステム制御装置から前記名前空間の更新に関する情報である名前空間更新情報を取得し、前記ストレージ装置におけるファイル識別情報とリンク情報に基づいて作成されたデータベースである名前空間複製データベースを、前記名前空間更新情報に基づいて更新する名前空間複製データベース更新部と、

前記名前空間複製データベース更新部による前記名前空間複製データベースへの反映が行われていない前記名前空間更新情報が喪失された場合、ファイル識別情報の更新時刻が

所定時刻以降のファイル識別情報である未更新ファイル識別情報と、該未更新ファイル識別情報に対応するリンク情報である未更新リンク情報とを、前記ファイルシステム制御装置から取得し、前記未更新ファイル識別情報と前記未更新リンク情報とに基づいて前記名前空間複製データベースを修正する名前空間複製データベース修正部と
を備える名前空間複製装置。
【請求項１０】
ストレージ装置上の名前空間の複製を行う名前空間複製方法であって、
前記ストレージ装置におけるファイル識別情報とリンク情報に基づいて作成されたデータベースである名前空間複製データベースを管理する名前空間複製装置において、前記名前空間の更新に関する情報である名前空間更新情報を、前記ストレージ装置の制御を行うファイルシステム制御装置から取得し、前記名前空間更新情報に基づいて前記名前空間複製データベースを更新する名前空間複製データベース更新ステップと、
前記名前空間複製データベース更新ステップによる前記名前空間複製データベースへの反映が行われていない前記名前空間更新情報が喪失された場合、前記名前空間複製装置において、ファイル識別情報の更新時刻が所定時刻以降のファイル識別情報である未更新ファイル識別情報と、該未更新ファイル識別情報に対応するリンク情報である未更新リンク情報とを、前記ファイルシステム制御装置から取得し、前記未更新ファイル識別情報と前記未更新リンク情報とに基づいて前記名前空間複製データベースを修正する名前空間複製データベース修正ステップと
を実行する名前空間複製方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、ストレージ装置上の名前空間の複製を行う、特に総なめ処理を行う際の性能改善を行う名前空間複製プログラム、名前空間複製装置、名前空間複製方法に関するものである。
【背景技術】
【０００２】
ＨＳＭ（Hierarchical Storage Management：階層記憶管理）は、テープライブラリなどの低速なストレージ装置（二次ストレージ）とハードディスクなどの高速なストレージ装置（一次ストレージ）を組み合わせることにより、安価な大容量ファイルシステムを構築するものである。
【０００３】
ＨＳＭ制御装置においては、一次ストレージにおいて長時間アクセスされていないファイルを特定し、そのファイルを二次ストレージに書き出し、アクセスが要求された時点で一次ストレージに移動することが必要となる。従来、これを実現するために、従来のＨＳＭ制御装置は、階層構造を持つファイルシステムの名前空間を総なめし、ファイルシステムがファイル単位に保持するアクセス時刻を参照することにより、二次ストレージに書き出すファイルを特定する方式を用いている。
【０００４】
なお、本発明の関連ある従来技術として、例えば、下記に示す特許文献１が知られている。このデータ処理装置は、メタデータデータの内容が更新されると、ログが採取され、このログを用いてファイルシステムの不整合の修正を行うものである。
【特許文献１】特開２０００－４８４９９５号公報
【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、上述した名前空間を総なめする方式のＨＳＭ制御装置には、以下の問題がある。
【０００６】

第１にファイルシステム総なめオーバヘッドの問題がある。従来のＨＳＭ制御装置では階層構造を持つファイル名前空間を定期的に総なめするために、オーバヘッドが大きくなってしまう。

【０００７】

第２に名前空間の排他問題がある。ＨＳＭ制御装置が名前空間を総なめしている間に、ｒｅｎａｍｅ操作などのファイル名変更操作が行われると、総なめの過程で求めたパス名が、実際には存在しない不当なものとなってしまう。このため、ＨＳＭ制御装置は、顧客が設定したポリシと矛盾するデータ移動操作を行ってしまう可能性がある。例えば、検索の途中で、上位ディレクトリがゴミ箱に移されたとすると、ゴミ箱全体を移動対象としてしまうようなことが起こる。こうした問題を防ごうとすると、ＨＳＭ制御装置は総なめの過程で、頻繁に矛盾をチェックし、矛盾があれば総なめをやり直すことが必要となり、論理が非常に複雑となるとともにオーバヘッドが大幅に増加する。

【０００８】

第３にＨＳＭポリシ制御の柔軟性がある。一般に階層構造の名前空間は格納されているファイル群の性格を表しているため、ＨＳＭポリシも名前空間に基づいて設定する（あるディレクトリ以下の全ファイルなど）のが自然である。しかし、上述した名前空間の排他問題により、名前空間に基づく複雑なポリシ制御を実現することが難しいという問題があった。

【０００９】

第４に二次ストレージに退避されたデータの属性情報不足の問題がある。また上述した名前空間の排他問題により、二次ストレージに格納されるデータに正しいパス名を付加することが難しい。このため、二次ストレージに格納されたデータはファイルシステムのメタデータのみからしかアクセスできないことになり、ファイルシステムのメタデータが壊れると、二次ストレージ上にデータは残っているのにもかかわらず、パス名と対応づけることができないため、ファイルデータを復旧することができないという問題があった。

【００１０】

本発明は上述した問題点を解決するためになされたものであり、ストレージ装置上の名前空間の複製を効率的に行う名前空間複製プログラム、名前空間複製装置、名前空間複製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】

【００１１】

上述した課題を解決するため、本発明は、ストレージ装置上の名前空間の複製をコンピュータに実行させる名前空間複製プログラムであって、前記ストレージ装置の制御を行うファイルシステム制御装置から前記名前空間の更新に関する情報である名前空間更新情報を取得し、前記ストレージ装置におけるファイル識別情報とリンク情報に基づいて作成されたデータベースである名前空間複製データベースを、前記名前空間更新情報に基づいて更新する名前空間複製データベース更新ステップと、前記名前空間複製データベース更新ステップによる前記名前空間複製データベースへの反映が行われていない前記名前空間更新情報が喪失された場合、ファイル識別情報の更新時刻が所定時刻以降のファイル識別情報である未更新ファイル識別情報と、該未更新ファイル識別情報に対応するリンク情報である未更新リンク情報とを、前記ファイルシステム制御装置から取得し、前記未更新ファイル識別情報と前記未更新リンク情報とに基づいて前記名前空間複製データベースを修正する名前空間複製データベース修正ステップとをコンピュータに実行させるものである。

【００１２】

また、本発明に係る名前空間複製プログラムにおいて、前記名前空間更新情報は、名前空間の更新の内容である名前空間更新内容と該更新の時刻である名前空間更新時刻を含み、前記名前空間複製データベース修正ステップは、前記名前空間複製データベース更新ステップにより前記名前空間複製データベースに反映された名前空間更新情報に含まれる名前空間更新時刻のうち、最終のものを前記所定時刻とすることを特徴とするものである。

【００１３】

また、本発明に係る名前空間複製プログラムにおいて、前記名前空間複製データベース修正ステップは、前記未更新ファイル識別情報が示すファイルのうちディレクトリファイルの持つリンク情報を、前記未更新リンク情報とすることを特徴とするものである。

【００１４】

また、本発明に係る名前空間複製プログラムにおいて、前記名前空間複製データベース情報修正ステップは、前記所定時刻を前記ファイルシステム制御装置へ通知することにより、前記未更新ファイル識別情報と前記未更新リンク情報を抽出させ、取得することを特徴とするものである。

【００１５】

また、本発明に係る名前空間複製プログラムにおいて、前記名前空間複製データベース情報修正ステップは、前記名前空間複製データベースの修正が完了する前に、前記名前空間複製データベース更新ステップにより前記名前空間複製データベースに反映されていない名前空間更新情報を取得し、該名前空間更新情報が前記未更新ファイル識別情報に無関係である場合、取得した該名前空間更新情報に基づいて前記名前空間複製データベースを更新することを特徴とするものである。

【００１６】

また、本発明に係る名前空間複製プログラムにおいて、１つの前記リンク情報は、１つのディレクトリファイルのｉｎｏｄｅ情報と該ディレクトリファイルに含まれる子のファイルのｉｎｏｄｅ情報と該ディレクトリファイルに含まれる子のファイルの名前情報とを含み、前記名前空間複製データベースは、前記リンク情報毎のエントリを持つことを特徴とするものである。

【００１７】

また、本発明に係る名前空間複製プログラムにおいて、前記ファイル識別情報は、ｉｎｏｄｅ情報であり、前記リンク情報は、１つのディレクトリファイルのｉｎｏｄｅ番号と該ディレクトリファイルの子である１つのファイルのｉｎｏｄｅ番号とを含むことを特徴とするものである。

【００１８】

また、本発明に係る名前空間複製プログラムにおいて、前記名前空間複製データベース情報修正ステップは、前記名前空間複製データベースの修正が完了する前に、前記名前空間複製データベース更新ステップにより前記名前空間複製データベースに反映されていない名前空間更新情報を取得し、該名前空間更新情報が前記未更新ファイル識別情報に関係する場合、該名前空間更新情報の名前空間変更時刻と該名前空間更新情報に関係する前記未更新ファイル識別情報の更新時刻とを比較することにより、該名前空間更新情報と前記未更新ファイル識別情報のうち新しい方に基づいて前記名前空間複製データベースを修正することを特徴とするものである。

【００１９】

また、本発明に係る名前空間複製プログラムにおいて、前記名前空間更新情報は、前記ファイルシステム制御装置により所定間隔毎にまとめて送信され、前記名前空間複製データベース情報更新ステップは、前記名前空間更新情報を取得する度に、該名前空間更新情報に基づいて前記名前空間複製データベースを更新することを特徴とするものである。

【００２０】

また、本発明は、ストレージ装置上の名前空間の複製を行う名前空間複製装置であって、前記ストレージ装置の制御を行うファイルシステム制御装置から前記名前空間の更新に関する情報である名前空間更新情報を取得し、前記ストレージ装置におけるファイル識別情報とリンク情報に基づいて作成されたデータベースである名前空間複製データベースを、前記名前空間更新情報に基づいて更新する名前空間複製データベース更新部と、前記名前空間複製データベース更新部による前記名前空間複製データベースへの反映が行われていない前記名前空間更新情報が喪失された場合、ファイル識別情報の更新時刻が所定時刻以降のファイル識別情報である未更新ファイル識別情報と、該未更新ファイル識別情報に対応するリンク情報である未更新リンク情報とを、前記ファイルシステム制御装置から取

得し、前記未更新ファイル識別情報と前記未更新リンク情報とに基づいて前記名前空間複製データベースを修正する名前空間複製データベース修正部とを備えたものである。

【００２１】

また、本発明に係る名前空間複製装置において、前記名前空間更新情報は、名前空間の更新の内容である名前空間更新内容と該更新の時刻である名前空間更新時刻を含み、前記名前空間複製データベース修正部は、前記名前空間複製データベース更新部により前記名前空間複製データベースに反映された名前空間更新情報に含まれる名前空間更新時刻のうち、最終のものを前記所定時刻とすることを特徴とするものである。

【００２２】

また、本発明は、ストレージ装置上の名前空間の複製を行う名前空間複製方法であって、前記ストレージ装置におけるファイル識別情報とリンク情報に基づいて作成されたデータベースである名前空間複製データベースを管理する名前空間複製装置において、前記名前空間の更新に関する情報である名前空間更新情報を、前記ストレージ装置の制御を行うファイルシステム制御装置から取得し、前記名前空間更新情報に基づいて前記名前空間複製データベースを更新する名前空間複製データベース更新ステップと、前記名前空間複製データベース更新ステップによる前記名前空間複製データベースへの反映が行われていない前記名前空間更新情報が喪失された場合、前記名前空間複製装置において、ファイル識別情報の更新時刻が所定時刻以降のファイル識別情報である未更新ファイル識別情報と、該未更新ファイル識別情報に対応するリンク情報である未更新リンク情報とを、前記ファイルシステム制御装置から取得し、前記未更新ファイル識別情報と前記未更新リンク情報とに基づいて前記名前空間複製データベースを修正する名前空間複製データベース修正ステップとを実行するものである。

【００２３】

また、本発明に係る名前空間複製方法において、前記名前空間更新情報は、名前空間の更新の内容である名前空間更新内容と該更新の時刻である名前空間更新時刻を含み、前記名前空間複製データベース修正ステップは、前記名前空間複製データベース更新ステップにより前記名前空間複製データベースに反映された名前空間更新情報に含まれる名前空間更新時刻のうち、最終のものを前記所定時刻とすることを特徴とするものである。

【００２４】

また、本発明に係る名前空間複製方法において、前記名前空間複製データベース修正ステップは、前記未更新ファイル識別情報が示すファイルのうちディレクトリファイルの持つリンク情報を、前記未更新リンク情報とすることを特徴とするものである。

【００２５】

また、本発明に係る名前空間複製方法において、前記名前空間複製データベース情報修正ステップは、前記名前空間複製装置において、前記所定時刻を前記ファイルシステム制御装置に通知し、前記ファイルシステム制御装置において、ファイル識別情報の更新時刻が前記所定時刻以降のファイル識別情報を列挙し、該ファイル識別情報を未更新ファイル識別情報として前記名前空間複製装置へ送信することにより、前記名前空間複製装置において、前記未更新ファイル識別情報を取得することを特徴とするものである。

【００２６】

また、本発明に係る名前空間複製方法において、前記名前空間複製データベース情報修正ステップは、前記名前空間複製装置において、前記名前空間複製データベースの修正が完了する前に、前記名前空間複製データベース更新ステップにより前記名前空間複製データベースに反映されていない名前空間更新情報を取得し、該名前空間更新情報が前記未更新ファイル識別情報に無関係である場合、取得した該名前空間更新情報に基づいて前記名前空間複製データベースを更新することを特徴とするものである。

【００２７】

また、本発明に係る名前空間複製方法において、１つの前記リンク情報は、１つのディレクトリファイルのｉｎｏｄｅ情報と該ディレクトリファイルに含まれる子のファイルのｉｎｏｄｅ情報と該ディレクトリファイルに含まれる子のファイルの名前情報とを含み、

前記名前空間複製データベースは、前記リンク情報毎のエントリを持つことを特徴とするものである。

【００２８】

また、本発明に係る名前空間複製方法において、前記ファイル識別情報は、ｉｎｏｄｅ情報であり、前記リンク情報は、１つのディレクトリファイルのｉｎｏｄｅ番号と該ディレクトリファイルの子である１つのファイルのｉｎｏｄｅ番号とを含むことを特徴とするものである。

【００２９】

また、本発明に係る名前空間複製方法において、前記名前空間複製データベース情報修正ステップは、前記名前空間複製装置において、前記名前空間複製データベースの修正が完了する前に、前記名前空間複製データベース更新ステップにより前記名前空間複製データベースに反映されていない名前空間更新情報を取得し、該名前空間更新情報が前記未更新ファイル識別情報に関係する場合、該名前空間更新情報の名前空間変更時刻と該名前空間更新情報に関係する前記未更新ファイル識別情報の更新時刻とを比較することにより、該名前空間更新情報と前記未更新ファイル識別情報のうち新しい方に基づいて前記名前空間複製データベースを修正することを特徴とするものである。

【００３０】

また、本発明に係る名前空間複製方法において、前記名前空間複製データベース情報更新ステップは、前記ファイルシステム制御装置において、整然停止を行う場合、前記名前空間複製データベースの維持を指示する情報であるデータベース維持情報を前記ストレージ装置に記録し、前記ファイルシステム制御装置の起動時に前記ストレージ装置に前記データベース維持情報がない場合、前記名前空間複製データベース更新ステップによる前記名前空間複製データベースへの反映が行われていない前記名前空間更新情報が喪失されたと判断し、前記名前空間複製装置に名前空間複製データベース修正ステップを実行させることを特徴とするものである。

【図面の簡単な説明】

【００３１】

【図１】前提技術１に係るＨＳＭ装置の構成の一例を示すブロック図である。

【図２】前提技術１に係るファイル情報取得処理の動作の一例を示すフローチャートである。

【図３】前提技術１に係る名前空間におけるディレクトリの階層構造の一例を示す図である。

【図４】前提技術１に係るファイル情報取得処理の動作の一例を示すフローチャートである。

【図５】前提技術１に係るイベントデータ反映処理の動作の一例を示すフローチャートである。

【図６】前提技術１に係るマイグレート決定処理の動作の一例を示すフローチャートである。

【図７】実施の形態１に係るＨＳＭシステムの構成の一例を示すブロック図である。

【図８】実施の形態１に係るＨＳＭシステムの詳細な構成と動作の一例を示すブロック図である。

【図９】実施の形態１に係る名前空間複製モード決定処理の動作の一例を示すフローチャートである。

【図１０】実施の形態１に係る名前空間に関するデータ構造の一例を示すブロック図である。

【図１１】実施の形態１に係るイベントの種類と内容の一例を示す表である。

【図１２】実施の形態１に係る名前空間ＤＢ修正処理の動作の一例を示すシーケンス図である。

【図１３】イベント喪失が発生した時点における一次ストレージの名前空間の内容の一例を示すツリー構造図である。

【図１４】イベント喪失が発生した時点における名前空間テーブルの内容の一例を示すツリー構造図である。
【図１５】ｉｎｏｄｅ情報が修正された時点における名前空間テーブルの内容の一例を示すツリー構造図である。
【図１６】修正されたｉｎｏｄｅ情報と無関係のイベントが反映された時点における名前空間テーブルの内容の一例を示すツリー構造図である。
【図１７】リンク情報が修正された時点における名前空間テーブルの内容の一例を示すツリー構造図である。
【発明を実施するための最良の形態】
【００３２】
以下、本発明の前提技術について図面を参照しつつ説明する。
【００３３】
前提技術１.
前提技術１においては、ＨＳＭ制御装置であるサーバについて説明する。
【００３４】
まず、前提技術１に係るサーバを有するＨＳＭ装置の構成について説明する。
【００３５】
図１は、前提技術１に係るＨＳＭ装置の構成の一例を示すブロック図である。最近アクセスされたファイルを格納しているディスク装置などの高速ストレージ装置である一次ストレージ１、および長時間アクセスされていないファイルデータが格納されるテープライブラリ装置などの低速ストレージ装置である二次ストレージ２と、前提技術１に係るＨＳＭ制御装置であり、ファイルデータをアクセスするアプリケーションが動作するサーバ３から構成される。
【００３６】
また、サーバ３は、アプリケーション部１１、ファイルシステム制御部１２、名前空間複製部１３、名前空間追随部１４、名前空間複製ＤＢ（Database）１５、マイグレート決定部１６を備える。また、ファイルシステム制御部１２は、イベントデータ記録部２１を備える。
【００３７】
次に、サーバ３の各部について説明する。
【００３８】
イベントデータ記録部２１は、アプリケーションプログラムが発行したファイル操作要求の履歴をイベントデータとして蓄積するファイルシステム制御部１２内に配置されるプログラムである。イベントデータ記録部２１は、アプリケーション部１１が発行したファイル操作要求の内容をイベントデータに変換してメモリ上に蓄積しておき、一定量たまったところで名前空間複製部１３や名前空間追随部１４に渡す。イベントデータの受け渡しは、通信を使用してもよいし、専用のファイルを介して受け渡してもよい。
【００３９】
名前空間複製部１３は、アプリケーション部１１の動作と平行して、ファイルシステムの名前空間の複製を行うプログラムである。名前空間複製部１３は、ファイルシステムの名前空間をたどり、存在するファイルのファイル情報を取得する。このファイル情報と、ファイル情報取得中にイベントデータ記録部２１から受け取ったイベントデータを組み合わせて、名前空間の初期複製を名前空間複製ＤＢ１５として完成させる。
【００４０】
名前空間追随部１４は、名前空間の初期複製が完成した後、イベントデータ記録部２１から受け取ったイベントデータに従って複製を更新し、名前空間複製ＤＢ１５を最新の状態に維持する機能を受け持つ。また、名前空間追随部１４は、通知されたファイルアクセスやアーカイブ状態を名前空間複製ＤＢ１５に反映する役割も担う。
【００４１】
マイグレート決定部１６は、ポリシ制御の一例として、名前空間複製部１３が設定した

ファイルアクセス記録とユーザが設定したポリシに従い、一次ストレージ１において長時間アクセスされていないファイルを二次ストレージ２に追い出すため、ファイルシステム制御部１２に指示を出すプログラムである。通常、二次ストレージ２に追い出された（マイグレートされた）ファイルは、アプリケーション部１１がそのファイルをアクセスしたときに、ファイルシステム制御部１２が二次ストレージ２から一次ストレージ１に戻す（リコール）。また、ファイルを更新したタイミングで、ファイルシステム制御部１２により二次ストレージ２上のデータ（アーカイブデータ）が無効化される。二次ストレージ２上のデータはこのタイミングでは消えず、二次ストレージ２が不足するまで、バックアップデータとして残され、ファイルシステム障害時などのリカバリで使われる。

【００４２】

次に、イベントデータ、ファイル情報、名前空間複製ＤＢ１５の詳細について説明する。

【００４３】

まず、イベントデータについて説明する。

【００４４】

イベントデータ記録部２１により作成されるイベントデータ（ｅｖｅｎｔ）はファイルやディレクトリの生成や削除、ファイル名の変更、ファイルアクセス、アーカイブ状態変化などのファイル操作の内容を表しており、操作名と操作が行われた時刻に加え、それぞれ以下のデータを含む。ここで、アーカイブ状態変化とは、アーカイブデータの有効化・無効化、マイグレート、リコールなどの事象を含む。

【００４５】

（１） ファイルあるいはディレクトリの作成
ｅｖｅｎｔ．ｒｅｃｔｙｐｅ ＝ ｃｒｅａｔｅ
ｅｖｅｎｔ．ｍ＿ｉｎｏｄｅ＃ ＝ 親ディレクトリのｉｎｏｄｅ番号
ｅｖｅｎｔ．ｆｔｙｐｅ ＝
ｄｉｒ （ｍｋｄｉｒ時）あるいはｆｉｌｅ （ｃｒｅａｔｅ時）
ｅｖｅｎｔ．ｆｎａｍｅ ＝ 作成されたファイルの名前
ｅｖｅｎｔ．ｉｎｏｄｅ＃ ＝
作成されたファイルあるいはディレクトリのｉｎｏｄｅ番号
ｅｖｅｎｔ．ｔｉｍｅ ＝ このイベントが発生した時刻

【００４６】

（２） ファイルあるいはディレクトリの削除
ｅｖｅｎｔ．ｒｅｃｔｙｐｅ ＝ ｄｅｌｅｔｅ
ｅｖｅｎｔ．ｍ＿ｉｎｏｄｅ＃ ＝ 親ディレクトリのｉｎｏｄｅ番号
ｅｖｅｎｔ．ｆｔｙｐｅ ＝ ｄｉｒ （ｒｍｄｉｒ時）あるいはｆｉｌｅ （ｒｅｍｏｖｅ時）
ｅｖｅｎｔ．ｉｎｏｄｅ＃ ＝ 削除されたファイルあるいはディレクトリのｉｎｏｄｅ番号
ｅｖｅｎｔ．ｔｉｍｅ ＝ このイベントが発生した時刻

【００４７】

（３） ファイル名の変更
ｅｖｅｎｔ．ｒｅｃｔｙｐｅ ＝ ｒｅｎａｍｅ
ｅｖｅｎｔ．ｍ＿ｉｎｏｄｅ＃ ＝ 親ディレクトリのｉｎｏｄｅ番号
ｅｖｅｎｔ．ｆｔｙｐｅ ＝
ｄｉｒ（対象がディレクトリの場合）
あるいはｆｉｌｅ（対象がファイルの場合）
ｅｖｅｎｔ．ｉｎｏｄｅ＃ ＝
対象のファイルあるいはディレクトリのｉｎｏｄｅ番号
ｅｖｅｎｔ．ｔａｒｇｅｔ．ｍ＿ｉｎｏｄｅ＃ ＝
移動先ディレクトリのｉｎｏｄｅ番号

ｅｖｅｎｔ．ｔａｒｇｅｔ．ｆｎａｍｅ ＝
変更後のファイルあるいはディレクトリ名
ｅｖｅｎｔ．ｔｉｍｅ ＝ このイベントが発生した時刻
【００４８】
（４） ファイルアクセス（アプリケーションプログラムがファイルをｒｅａｄ／ｗｒｉｔｅ）
ｅｖｅｎｔ．ｒｅｃｔｙｐｅ ＝ ａｃｃｅｓｓ
ｅｖｅｎｔ．ｉｎｏｄｅ＃ ＝ ファイルのｉｎｏｄｅ番号
ｅｖｅｎｔ．ｔｉｍｅ ＝ このイベントが発生した時刻
【００４９】
（５） アーカイブ状態変化
ｅｖｅｎｔ．ｒｅｃｔｙｐｅ ＝ ａｒｃｈｉｖｅ
ｅｖｅｎｔ．ｉｎｏｄｅ＃ ＝ ファイルのｉｎｏｄｅ番号
ｅｖｅｎｔ．ｍｉｇｒａｔｅ ＝
オン（マイグレート状態となった）
あるいはオフ（リコールが起動され、マイグレート状態でなくなった）
ｅｖｅｎｔ．ａｒｃｈｉｖｅ ＝
オン（二次記憶へのファイルデータの書き出しが完了し、アーカイブデータが有効となった）
あるいはオフ（ファイルが更新された結果アーカイブデータが無効となった）
ｅｖｅｎｔ．ｔｉｍｅ ＝ このイベントが発生した時刻
【００５０】
次に、ファイル情報について説明する。
【００５１】
名前空間複製復元中にファイルシステムから取得するファイル情報（ｆｓｔａｔ）には、以下のものがある。
ｆｓｔａｔ．ｍ＿ｉｎｏｄｅ＃ ＝ 親ディレクトリのｉｎｏｄｅ番号
ｆｓｔａｔ．ｆｔｙｐｅ ＝ ｄｉｒ（対象がディレクトリの場合）
あるいはｆｉｌｅ （対象がファイルの場合）
ｆｓｔａｔ．ｆｎａｍｅ ＝ ディレクトリあるいはファイルの名前
ｆｓｔａｔ．ｉｎｏｄｅ＃ ＝
ファイルあるいはディレクトリのｉｎｏｄｅ番号
ｆｓｔａｔ．ａｒｃｈｉｖｅ ＝ オン（アーカイブデータが有効なとき）
あるいはオフ（アーカイブデータが無効なとき）
ｆｓｔａｔ．ｍｉｇｒａｔｅ ＝ オン（マイグレート状態のとき）
あるいはオフ（マイグレートされていないとき）
ｆｓｔａｔ．ａｔｉｍｅ ＝ ファイルを最後にアクセスした時刻
ｆｓｔａｔ．ｔｉｍｅ ＝ ファイル情報取得時刻
【００５２】
次に、名前空間複製ＤＢ１５の構成について説明する。
【００５３】
名前空間複製ＤＢ１５は、以下のカラム（ｄｂｅ）を持つ、ディレクトリに設定されているファイルあるいはディレクトリ要素ごとにタプルを持つリレーショナルＤＢである。
【００５４】
ｄｂｅ．ｍ＿ｉｎｏｄｅ＃ ＝ 親ディレクトリのｉｎｏｄｅ番号
ｄｂｅ．ｆｔｙｐｅ ＝ ｄｉｒ（このタプルがディレクトリをあらわすとき） あるいは ｆｉｌｅ （このタプルがファイルを表すとき）
ｄｂｅ．ｆｎａｍｅ ＝ ファイルあるいはディレクトリの名前
ｄｂｅ．ｉｎｏｄｅ＃ ＝ ファイルあるいはディレクトリのｉｎｏｄｅ番号
ｄｂｅ．ａｒｃｈｉｖｅ ＝ オン（アーカイブデータが有効なとき）

あるいはオフ（アーカイブデータが無効なとき）
ｄｂｅ．ｍｉｇｒａｔｅ　＝　オン（マイグレート状態のとき）
あるいはオフ（マイグレートされていないとき）
ｄｂｅ．ａｔｉｍｅ　＝　ファイルを最後にアクセスした時刻
ｄｂｅ．ａｃｔｉｖｅ　＝　オン（ファイル情報取得済みのとき）
あるいはオフ（まだファイル情報を取得していないとき）

【００５５】

次に、サーバ３の動作について説明する。

【００５６】

図２は、前提技術１に係るファイル情報取得処理の動作の一例を示すフローチャートである。サーバ３は、名前空間複製処理（Ｓ１１）、名前空間追随処理（Ｓ１２）、マイグレート処理（Ｓ１３）を実行する。

【００５７】

次に、サーバ３における動作の詳細について説明する。

【００５８】

まず、名前空間複製処理について説明する。

【００５９】

名前空間複製処理は、名前空間複製部１３が名前空間の初期複製を作成する処理であり、ファイル情報取得処理とイベントデータ反映処理からなる。また、名前空間複製処理は、障害発生後のサーバ再立ち上げ時など、メモリ上に蓄積されていたイベントデータが失われ、名前空間複製ＤＢ１５の内容がファイルシステムの最新状態を反映できなくなったときに、名前空間複製ＤＢ１５を再作成する目的で動作する。このように名前空間複製ＤＢ１５を動的に再作成する構成では、イベントデータをイベント発生時に不揮発化する必要がなく、小さい容量のメモリに蓄積するのみで良く、後の名前空間複製ＤＢの追随のオーバヘッドを削減することができる。

【００６０】

名前空間複製部１３は、まず、ファイル情報取得処理として、親ディレクトリをオープンし、子ファイル名あるいは子ディレクトリ名を引数として指定し、ファイルシステムの情報取得機能（ｇｅｔｉｎｆｏ）を発行することにより求める。また、名前空間複製部１３は、パス名昇順（あるいは降順）とした名前空間をたどることにより、ファイルシステム内に存在するディレクトリ、ファイルの情報を漏れなく求める。この過程で見逃したものは、イベントデータとして記録されているので、後で補正する。

【００６１】

図３は、名前空間におけるディレクトリの階層構造の一例を示す図である。この名前空間は、ディレクトリの階層構造を持ち、ディレクトリ名やファイル名を昇順に左から右へソートしたものである。図４は、前提技術１に係るファイル情報取得処理の動作の一例を示すフローチャートである。

【００６２】

まず、名前空間複製部１３は、対象ファイルシステムのルートディレクトリを基点とし、ディレクトリを左下方向（ディレクトリ名の昇順）に順にたどり、最も左下のディレクトリを見つける。見つけた最も左下のディレクトリを対象ディレクトリとし、検索の過程で求めた対象ディレクトリのパス名を対象ディレクトリパス名とする（Ｓ２０１）。次に、名前空間複製部１３は、対象ディレクトリのファイル情報および対象ディレクトリ内に存在する全ファイルのファイル情報をファイル名昇順にひとつずつ順に求め、ファイル情報記録ファイルの末尾に順に書き込む（Ｓ２０２）。次に、名前空間複製部１３は、対象ディレクトリがルートディレクトリであるか否かの判断を行う（Ｓ２０３）。対象ディレクトリがルートディレクトリである場合（Ｓ２０３，Ｙ）、全ファイルを処理し終わったことを意味するのでファイル情報取得処理を終了する。

【００６３】

一方、対象ディレクトリがルートディレクトリでない場合（Ｓ２０３，Ｎ）、名前空間

複製部１３は、対象ディレクトリパス名から、対象ディレクトリのひとつ上のディレクトリパス名を求める、すなわち、パス名を構成する最終構成ディレクトリ名を取り除いたパス名を新しいパス名とする。次に、名前空間複製部１３は、求めたディレクトリパス名をルートディレクトリから下方に順に再度検索し、この検索で存在を確認できた最終ディレクトリを基点ディレクトリとする（Ｓ２０５）。パスの途中のディレクトリがｒｅｎａｍｅなどで名前空間の別の位置に動かされている場合、途中でみつからなくなるが、この部分は後続のファイル情報取得処理で見つかるか、イベントデータで必ず通知され、後で補正されるため、無視しても問題ない。

【００６４】

次に、名前空間複製部１３は、基点ディレクトリの内容を読み込み、基点ディレクトリ内に未処理のディレクトリがあるか否かの判断を行う（Ｓ２０６）。未処理のディレクトリがある場合（Ｓ２０６，Ｙ）、名前空間複製部１３は、未処理の最も左下のディレクトリを求め、これを対象ディレクトリとし（Ｓ２０７）、処理Ｓ２０２に移行する。未処理のディレクトリが存在しない、すなわち基点ディレクトリ内に対象ディレクトリパス名で示されるより大きなファイル名をもつディレクトリが存在しない場合（Ｓ２０６，Ｎ）、対象ディレクトリパス名を基点ディレクトリのパス名に設定し（Ｓ２０８）、処理Ｓ２０３に移行する。

【００６５】

次に、名前空間複製部１３は、対象ファイルシステムのファイル情報取得処理が全て終了すると、その間に発生したイベントデータをファイル情報に反映するイベントデータ反映処理を行う。ファイル情報記録ファイルを先頭から順によみ、ファイル情報記録ファイルに記録されている全てのファイル情報を処理したら、イベントデータ反映処理は終了する。

【００６６】

図５は、前提技術１に係るイベントデータ反映処理の動作の一例を示すフローチャートである。まず、名前空間複製部１３は、未処理のファイル情報を取り出し（Ｓ３０２）、ファイル情報に設定されていた情報取得時刻以前の時刻を持つ、イベントデータを順に取り出し、名前空間複製ＤＢ１５に反映する（Ｓ３０３）。

【００６７】

ここで、名前空間複製ＤＢ１５への反映について、イベントデータが、削除系、生成系、ファイル名の変更、ファイルアクセス、アーカイブ状態変化のそれぞれの場合について説明する。

【００６８】

イベントデータが削除系（ファイル削除，ディレクトリ削除）の場合、名前空間複製部１３は、削除対象ファイルあるいはディレクトリが既に名前空間複製ＤＢ１５に登録済みなら削除する。そうでなければ何もしない。ここで、以下の条件を全て満たすエントリが存在する場合、登録済みとみなす。

【００６９】

```
dbe.inode#  ==  event.inode#
dbe.m_inode#  ==  event.m_inode#
dbe.fname  ==  event.fname
```

【００７０】

イベントデータが生成系（ファイル生成，ディレクトリ生成）の場合、名前空間複製部１３は、作成されたファイルあるいはディレクトリが名前空間複製ＤＢ１５に登録済みでなければ情報取得済みで登録する。登録済みならこのイベントデータを無視し、何もしない。ここで、以下の条件を全て満たすエントリが存在する場合、登録済みとみなす。

【００７１】

```
dbe.inode#  ==  event.inode#
dbe.m_inode#  ==  event.m_inode#
dbe.fname  ==  event.fname
```

【００７２】
未登録時の設定内容を以下に示す。
【００７３】
ｄｂｅ．ｍ＿ｉｎｏｄｅ＃　＝　ｅｖｅｎｔ．　ｍ＿ｉｎｏｄｅ＃
ｄｂｅ．ｆｔｙｐｅ　＝　ｅｖｅｎｔ．　ｆｔｙｐｅ
ｄｂｅ．ｆｎａｍｅ　＝　ｅｖｅｎｔ．ｆｎａｍｅ
ｄｂｅ．ｉｎｏｄｅ＃　＝　ｅｖｅｎｔ．ｉｎｏｄｅ＃
ｄｂｅ．ａｒｃｈｉｖｅ　＝　オフ
ｄｂｅ．ｍｉｇｒａｔｅ　＝　オフ
ｄｂｅ．ａｔｉｍｅ　＝　ｅｖｅｎｔ．ｔｉｍｅ
ｄｂｅ．ａｃｔｉｖｅ　＝　オン
【００７４】
イベントデータがファイル名の変更（ｅｖｅｎｔ．ｒｅｃｔｙｐｅ　＝＝　ｒｅｎａｍｅ）の場合、名前空間複製部１３は、改名後と同じ名前をもつファイルあるいはディレクトリがすでに登録されていた場合（ファイル名と親ｉｎｏｄｅ番号で評価）、そのエントリを名前空間複製ＤＢ１５から削除する。ここで、以下の条件を全て満たすエントリが存在する場合、登録済みとみなす。
【００７５】
ｄｂｅ．ｎａｍｅ　＝＝　ｅｖｅｎｔ．ｔａｒｇｅｔ．ｆｎａｍｅ
ｄｂｅ．ｍ＿ｉｎｏｄｅ＃　＝＝
　ｅｖｅｎｔ．ｔａｒｇｅｔ．ｍ＿ｉｎｏｄｅ＃
ｄｂｅ．ｆｎａｍｅ　＝＝　ｅｖｅｎｔ．ｔａｒｇｅｔ．ｆｎａｍｅ
【００７６】
ここで、対象ファイルが名前空間複製ＤＢ１５に既に登録されているならそのエントリの親情報とファイル名を変更する。ここで、以下の条件を全て満たすエントリが存在する場合、登録済みとみなす。
【００７７】
ｄｂｅ．ｉｎｏｄｅ＃　＝＝　ｅｖｅｎｔ．ｉｎｏｄｅ＃
ｄｂｅ．ｍ＿ｉｎｏｄｅ＃　＝＝　ｅｖｅｎｔ．ｍ＿ｉｎｏｄｅ＃
ｄｂｅ．ｆｎａｍｅ　＝＝　ｅｖｅｎｔ．ｆｎａｍｅ
【００７８】
このときの変更内容を以下に示す。
【００７９】
ｄｂｅ．ｍ＿ｉｎｏｄｅ＃　＝　ｅｖｅｎｔ．ｔａｒｇｅｔ．ｍ＿ｉｎｏｄｅ＃
ｄｂｅ．ｎａｍｅ　＝　ｅｖｅｎｔ．ｔａｒｇｅｔ．ｆｎａｍｅ
【００８０】
ここで、対象ファイルが未登録なら、改名後のファイルを名前空間複製ＤＢ１５に新しいエントリとして登録する。
【００８１】
ｄｂｅ．ｉｎｏｄｅ＃　＝　ｅｖｅｎｔ．ｉｎｏｄｅ＃
ｄｂｅ．ｍ＿ｉｎｏｄｅ＃　＝　ｅｖｅｎｔ．ｔａｒｇｅｔ．ｍ＿ｉｎｏｄｅ＃
ｄｂｅ．ｎａｍｅ　＝　ｅｖｅｎｔ．ｔａｒｇｅｔ．ｆｎａｍｅ
ｄｂｅ．ａｃｔｉｖｅ　＝　オフ
【００８２】
イベントデータがファイルアクセス（ｅｖｅｎｔ．ｒｅｃｔｙｐｅ　＝＝　ａｃｃｅｓｓ）の場合、名前空間複製部１３は、対象ｉｎｏｄｅが未登録ならこのイベントデータを無視する。登録されていたら、登録済みのすべてのエントリのファイル最終アクセス時刻、アーカイブ情報、リコール情報を更新（ハードリンクがあるため）する。ここで、以下の条件を全て満たすエントリが存在する場合、登録済みとみなす。
【００８３】

ｄｂｅ．ｉｎｏｄｅ＃　＝＝　ｅｖｅｎｔ．ｉｎｏｄｅ＃

【００８４】

このときの変更内容を以下に示す。

【００８５】

ｄｂｅ．ａｔｉｍｅ　＝　ｅｖｅｎｔ．ｔｉｍｅ

【００８６】

イベントデータがアーカイブ状態変化（ｅｖｅｎｔ．ｒｅｃｔｙｐｅ　＝＝　ａｒｃｈｉｖｅ）の場合、対象ｉｎｏｄｅが未登録ならこのイベントデータを無視。登録されていたら、すべてのエントリのアーカイブ情報を更新（ハードリンクがあるため）する。ここで、以下の条件を全て満たすエントリが存在する場合、登録済みとみなす。

【００８７】

ｄｂｅ．ｉｎｏｄｅ＃　＝＝　ｅｖｅｎｔ．ｉｎｏｄｅ＃

【００８８】

このときの変更内容を以下に示す。

【００８９】

ｄｂｅ．ａｒｃｈｉｖｅ　＝　ｅｖｅｎｔ．ａｒｃｈｉｖｅ
ｄｂｅ．ｍｉｇｒａｔｅ　＝　ｅｖｅｎｔ．ｍｉｇｒａｔｅ

【００９０】

次に、名前空間複製部１３は、ファイル情報の内容を名前空間複製ＤＢ１５に未登録なら情報取得済みとして登録する（Ｓ３０５）。同一ｉｎｏｄｅ番号を持つタプルが登録されていた場合には、登録されている全てのエントリの内容を変更する。ここで、以下の条件をすべて満たすエントリが存在するとき、登録済みとみなす。

【００９１】

ｄｂｅ．ｉｎｏｄｅ＃　＝＝　ｆｓｔａｔ．ｉｎｏｄｅ＃
ｄｂｅ．ｆｎａｍｅ　＝＝　ｆｓｔａｔ．ｆｎａｍｅ
ｄｂｅ．ｍ＿ｉｎｏｄｅ＃　＝＝　ｆｓｔａｔ．ｍ＿ｉｎｏｄｅ＃

【００９２】

また、未登録時の設定内容を以下に示す。

【００９３】

ｄｂｅ．ｍ＿ｉｎｏｄｅ＃　＝　ｆｓｔａｔ．　ｍ＿ｉｎｏｄｅ＃
ｄｂｅ．ｆｔｙｐｅ　＝　ｆｓｔａｔ．　ｆｔｙｐｅ
ｄｂｅ．ｆｎａｍｅ　＝　ｆｓｔａｔ．ｆｎａｍｅ
ｄｂｅ．ｉｎｏｄｅ＃　＝　ｆｓｔａｔ．ｉｎｏｄｅ＃
ｄｂｅ．ａｒｃｈｉｖｅ　＝　ｆｓｔａｔ．ａｒｃｈｉｖｅ
ｄｂｅ．ｍｉｇｒａｔｅ　＝　ｆｓｔａｔ．ｍｉｇｒａｔｅ
ｄｂｅ．ａｔｉｍｅ　＝　ｆｓｔａｔ．ａｔｉｍｅ
ｄｂｅ．ａｃｔｉｖｅ　＝　オン

【００９４】

また、同一ｉｎｏｄｅ番号が既に登録済み（すなわちｄｂｅ．ｉｎｏｄｅ＃　＝　ｆｓｔａｔ．ｉｎｏｄｅ＃の場合）の設定内容を以下に示す。

【００９５】

ｄｂｅ．ａｒｃｈｉｖｅ　＝　ｆｓｔａｔ．ａｒｃｈｉｖｅ
ｄｂｅ．ｍｉｇｒａｔｅ　＝　ｆｓｔａｔ．ｍｉｇｒａｔｅ
ｄｂｅ．ａｔｉｍｅ　＝　ｆｓｔａｔ．ａｔｉｍｅ
ｄｂｅ．ａｃｔｉｖｅ　＝　オン

【００９６】

次に、名前空間複製部１３は、記録されていた全ファイル情報の処理を終了すると、名前空間の変更との競合のため情報取得で見逃した名前空間のセグメント（情報取得済みが表示されていないディレクトリ）が存在するか否かの判断を行う（Ｓ３１１）。存在しない場合（Ｓ３１１，Ｎ）、このフローを終了する。一方、存在する場合（Ｓ３１１，Ｙ）

、そのディレクトリをルートとするファイル情報取得処理、およびその間に発生したイベントデータ反映処理を行い（Ｓ３１２）、処理Ｓ３１１へ戻り、次の情報取得済みが表示されていないディレクトリを見つけ、この処理を繰り返す。

【００９７】

次に、名前空間追随処理について説明する。

【００９８】

名前空間追随部１４は、名前空間複製処理が完了した後に発生したイベントデータをイベントデータ記録部２１から受け取り、名前空間複製ＤＢ１５に順次反映していく。イベントデータ反映処理は名前空間複製処理とほぼ同じだが、ファイル情報を用いない分、単純となる。

【００９９】

イベントデータが削除系ファイル操作イベント（ファイル削除、ディレクトリ削除）である場合、名前空間追随部１４は、イベントデータで示されるｉｎｏｄｅ番号、親ｉｎｏｄｅ番号、ファイル名を全て含むエントリを名前空間複製ＤＢ１５上から削除する。

【０１００】

イベントデータが生成系ファイル操作イベント（ファイル生成、ディレクトリ生成）である場合、名前空間追随部１４は、イベントデータで示されるｉｎｏｄｅ番号を含むエントリを名前空間複製ＤＢ１５上に登録し、イベントデータで伝えられた属性（タイプ）、および親ｉｎｏｄｅ番号を設定する。

【０１０１】

イベントデータがファイル名の変更（ｒｅｎａｍｅ）でターゲットと同じファイルがあれば、名前空間追随部１４は削除する。また、名前空間追随部１４は、ソースの親属性を変更する。

【０１０２】

イベントデータがファイルアクセスイベントである場合、名前空間追随部１４は、イベントデータで伝えられたアクセス時刻をｉｎｏｄｅ番号で特定し、名前空間複製ＤＢ１５に設定する。

【０１０３】

イベントデータがアーカイブ状態変化である場合、名前空間追随部１４は、アーカイブ情報を更新する。

【０１０４】

次に、マイグレート処理について説明する。

【０１０５】

マイグレート決定部１６は、ファイルシステムが提供するコマンドなどを使い、一次ストレージ１の空きスペース状況を定期的に調べ、空きスペース量がユーザにより指定された量以下になった場合、名前空間複製ＤＢ１５に設定されている情報を使って、マイグレートの対象ファイルを決定し、ファイルシステム制御部１２にマイグレートを要求する。この際、マイグレート決定部１６は、名前空間複製ＤＢ１５から求めたファイルのパス名をファイルシステム制御部１２に渡し、ファイルデータとともに二次ストレージ２に書き出してもらう。マイグレート決定処理は、ユーザポリシに応じて様々な実装を行うことができるが、一例を以下に示す。

【０１０６】

図６は、前提技術１に係るマイグレート決定処理の動作の一例を示すフローチャートである。まず、マイグレート決定部１６は、一次ストレージ１の不足が深刻であるか否かの判断を行う（Ｓ４０１）。

【０１０７】

一次ストレージ１の不足が深刻である場合（Ｓ４０１，Ｙ）、マイグレート決定部１６は、名前空間複製ＤＢ１５を検索し、アーカイブ済みでかつマイグレート済みではないファイルを見つけ（Ｓ４１１）、見つけた全てのファイルに対し以下のリリース処理（一次ストレージ域の解放）を行う。次に、マイグレート決定部１６は、見つけたファイルのう

ち未処理のファイルがあるか否かの判断を行う（Ｓ４１２）。

【０１０８】

未処理のファイルがなければ（Ｓ４１２，Ｎ）、このフローを終了する。一方、未処理のファイルがあれば（Ｓ４１２，Ｙ）、マイグレート決定部１６は、名前空間複製ＤＢ１５に設定されているｉｎｏｄｅ番号を引数としてファイルシステム制御部１２に対象ファイルのリリース（一次ストレージ解放）を要求する（Ｓ４１３）。次に、マイグレート決定部１６は、ファイルシステム制御部１２からの応答を得ると、処理Ｓ４１２へ戻り、次の対象ファイルの処理を行う。

【０１０９】

ここで、マイグレート決定部１６は、名前空間複製ＤＢ１５はファイルシステムに遅れて追随するため、実際にはファイルが存在しなくなっている場合や、アーカイブが無効になっている場合があり、この場合にはファイルシステム制御部１２がエラー応答を返す。ファイルがアーカイブ済みであった場合、ファイルシステム制御部１２はそのファイルに割り当てていた一次ストレージ領域を解放して正常応答を返す。

【０１１０】

一方、一次ストレージ１が不足しているがそれほど深刻ではない場合（Ｓ４０１、Ｎ）、マイグレート決定部１６は、深刻な不足が発生したときに、事態をただちに改善できるようにするため、一定時間以上アクセスされていないファイルをアーカイブする。このため、マイグレート決定部１６は、名前空間複製ＤＢ１５を検索し、最終アクセス時刻が所定の時刻（例えば現時刻－１日）以前でかつ、アーカイブ無効（アーカイブ済みでない）なものを見つける（Ｓ４２１）。次に、マイグレート決定部１６は、見つけたファイルのうち未処理のファイルがあるか否かの判断を行う（Ｓ４２２）。

【０１１１】

未処理のファイルがなければ（Ｓ４２２，Ｎ）、このフローを終了する。一方、未処理のファイルがあれば（Ｓ４２２，Ｙ）、マイグレート決定部１６は、名前空間複製ＤＢ１５に設定されている親ｉｎｏｄｅ番号をキーとして、繰り返し名前空間複製ＤＢ１５を検索することで、ファイルのパス名を求める（Ｓ４２３）。次に、マイグレート決定部１６は、ｉｎｏｄｅ番号、ファイルパス名を引数としたアーカイブ要求をファイルシステム制御部１２に出す（Ｓ４２４）。ここで、ファイルシステム制御部１２は、指定されたファイルのデータとファイルパス名、ｉｎｏｄｅ番号を一括して二次ストレージ上に書き出し、処理Ｓ４２２へ戻り、次の対象ファイルの処理を行う。ここで、要求されたファイルが存在しなくなっている場合、ファイルシステム制御部１２はエラーを応答し、要求を無視する。

【０１１２】

次に、その他の各部の動作について説明する。

【０１１３】

まず、ファイルシステム制御部１２について説明する。

【０１１４】

まず、マイグレート決定部１６からのリリース要求があった場合、ファイルシステム制御部１２は、リリース要求を処理し、二次ストレージにファイルデータのコピーが存在する（アーカイブ済み）なら、一次ストレージを返却し、マイグレート済みとする。このとき、イベントデータ記録部２１はアーカイブ状態変化イベントを作成する。

【０１１５】

ｅｖｅｎｔ．ｒｅｃｔｙｐｅ　＝　ａｒｃｈｉｖｅ
ｅｖｅｎｔ．ａｒｃｈｉｖｅ　＝　オン
ｅｖｅｎｔ．ｍｉｇｒａｔｅ　＝　オン

【０１１６】

また、マイグレート決定部１６からのアーカイブ要求があった場合、ファイルシステム制御部１２は、アーカイブ要求を処理し、ファイルデータの二次ストレージ２への書き出しを起動し、マイグレート決定部１６に復帰する。この際、二次ストレージ２に書き出す

データのヘッダ部にファイルのマイグレート決定部１６から通知されたファイルのパス名を付加して書き出す。二次ストレージ２への書き出しが完了すると、イベントデータ記録部２１はアーカイブ状態変化イベントを作成する。

【０１１７】

ｅｖｅｎｔ．ｒｅｃｔｙｐｅ　＝　ａｒｃｈｉｖｅ
ｅｖｅｎｔ．ａｒｃｈｉｖｅ　＝　オン
ｅｖｅｎｔ．ｍｉｇｒａｔｅ　＝　オフ

【０１１８】

また、アプリケーション部１１がマイグレート済みファイルをアクセスしようとした場合、ファイルシステム制御部１２は、アプリケーション部１１がアクセスしようとしたタイミングで、一次ストレージ１上に領域を新たに割り当て、二次ストレージ２上のデータをその領域に読み込む。その後、イベントデータ記録部２１は、リコール完了を表示したアーカイブ状態変化イベントを作成する。

【０１１９】

ｅｖｅｎｔ．ｒｅｃｔｙｐｅ　＝　ａｒｃｈｉｖｅ
ｅｖｅｎｔ．ａｒｃｈｉｖｅ　＝　オン
ｅｖｅｎｔ．ｍｉｇｒａｔｅ　＝　オフ

【０１２０】

また、アプリケーション部１１がファイル操作（ファイル生成・削除、ディレクトリ生成・削除、ファイルｒｅａｄ／ｗｒｉｔｅ）を要求した場合、ファイルシステム制御部１２は要求を処理し、正常に完了した時点で、イベントデータ記録部２１は対応するイベントデータを作成する。

【０１２１】

名前空間複製部１３からｇｅｔｉｎｆｏでファイル情報を要求された場合、ファイルシステム制御部１２は、指定されたファイルが親ディレクトリに存在することを確認した上で、指定されたファイルのファイル情報を返す。存在しなければ、エラーを応答。エラーが返された場合の名前空間複製部１３はそのファイルがなかったものとして処理を続ける。

【０１２２】

次に、イベントデータ記録部２１について説明する。

【０１２３】

イベントデータ記録部２１は、ファイルシステム制御部１２内に存在し、ファイルシステム制御部１２の説明で述べたタイミングでイベントデータを作成し、メモリ上に蓄積する部分である。また、イベントデータ記録部２１は、メモリ上に蓄積されたイベントデータが一定量以上となった、あるいは最後に通知してから、一定時間経過したときに、メモリ上に蓄積されていたイベントデータを一括して、名前空間追随部１４あるいは名前空間複製部１３に通知する。また、システム停止時にも、イベントデータ記録部２１が蓄積していたイベントデータを名前空間追随部１４に通知し、名前空間追随部１４がメモリ上に蓄積されているイベントデータを名前空間複製ＤＢ１５に全て反映する、システム停止処理を行う。

【０１２４】

また、イベントデータ記録部２１では、通知するデータ量を削減するため、以下の最適化を施す。まず、イベントデータ記録部２１がファイルアクセスイベントを作成する場合、メモリ上に蓄積されている未通知のイベントデータの中に同じファイルに対するファイルアクセスイベントが含まれているなら、後続のファイルアクセスイベントは捨てる。すなわち、メモリ上に蓄積しない。また、イベントデータ記録部２１がファイル削除イベントの作成を依頼されたときに、対応するファイル生成イベントが未通知のイベントデータとして含まれるなら、ファイル生成イベントをメモリ上で無効化し、イベントデータ通知の対象から取り除く。

【０１２５】

次に、サーバ３におけるシステム立ち上げの処理について説明する。

【０１２６】

システムを正常終了した場合、上述したように名前空間追随部１４がメモリ上に滞留していたイベントデータを一括して名前空間複製ＤＢ１５に反映する正常終了処理を行うため、次回立ち上げ時に名前空間複製部１３を動作させる必要はない。一方、障害発生の場合、その後の再立ち上げ時には、名前空間複製部１３を動作させ、名前空間複製ＤＢ１５を再初期化するシステム異常終了後起動処理を行う。なお、この場合でも、障害発生直前の名前空間情報は残っているので、名前空間複製の再初期化が完了するまでの間にマイグレーション対象を決定する必要が発生した場合には、マイグレート決定部は古い複製を使って処理を行う。

【０１２７】

なお、前提技術１においては、名前空間複製ＤＢ１５に基づくポリシ制御の例としてマイグレート決定部１６について説明したが、ＨＳＭ制御における他のポリシ制御を名前空間複製ＤＢ１５に基づいて行う構成としても良い。

【０１２８】

前提技術１において、名前空間が更新される度にファイルシステム制御部１２から名前空間追随部１４へのイベント通知を行うとファイルシステム制御部１２に負荷が掛かり過ぎるため、ファイルシステム制御部１２はイベントをある程度溜めてからまとめてイベント通知を行う。しかしながら、通信障害やファイルシステム制御部１２のクラッシュなどの原因によりファイルシステム制御部１２に滞留していたイベントの喪失が発生した場合、名前空間複製部１３は、それまでの名前空間複製ＤＢ１５の内容を一旦破棄し、一次ストレージ１の名前空間を全てスキャンする総なめ処理を行い、複製を一から作り直す。ここで、喪失したイベント数が少ないとしても、総なめ処理の負荷は大きい。

【０１２９】

総なめ処理は、深いところから順に名前空間のスキャンを行う。また、名前空間複製ＤＢ復旧処理の途中、名前空間の更新のイベントが通知されると、名前空間複製部１３は、名前空間において更新箇所を含むツリーを再スキャンする。従って、名前空間の頻繁な更新が行われると、総なめ処理の収束が遅延する。特にファイルシステムが巨大である場合、名前空間複製ＤＢ復旧処理が終わらない場合がある。

【０１３０】

以下、本発明の実施の形態について図面を参照しつつ説明する。

【０１３１】

実施の形態１．

本実施の形態では、前提技術１と同様にして名前空間複製ＤＢを作成、更新するＨＳＭシステムにおいて、ＦＳ（File System）制御サーバ（ファイルシステム制御装置）からのイベントが喪失された場合、効率的に名前空間複製ＤＢの修正を行うＨＳＭシステムについて述べる。

【０１３２】

まず、本実施の形態に係るＨＳＭシステムの構成について説明する。

【０１３３】

図７は、本実施の形態に係るＨＳＭシステムの構成の一例を示すブロック図である。このＨＳＭシステムは、ユーザアプリケーション１１１、ＦＳ制御サーバ１１２、ストレージ管理サーバ１３１、名前空間複製ＤＢ１３２、一次ストレージ１３３、二次ストレージ１３４を備える。ユーザアプリケーション１１１とＦＳ制御サーバ１１２は、ＬＡＮ（Local Area Network）１１３ａで接続されている。ＦＳ制御サーバ１１２とストレージ管理サーバ１３１は接続されている。また、ＦＳ制御サーバ１１２とストレージ管理サーバ１３１と一次ストレージ１３３は、ＳＡＮ（Storage Area Network）１１４ａで接続されており、ストレージ管理サーバ１３１と二次ストレージ１３４と名前空間複製ＤＢ１３２は、ＳＡＮ１１４ｂで接続されている。

【０１３４】

図８は、本実施の形態に係るＨＳＭシステムの詳細な構成と動作の一例を示すブロック図である。ここで、ＦＳ制御サーバ１１２は、ＡＣ（Access Client）１２１、ＭＤＳ（Meta Data Server）１２２、ＨＳＭＡ（HSM Agent）１２３を備える。また、ＭＤＳ１２２は、イベントキュー１２４を備える。なお、本実施の形態における一次ストレージ１３３は、前提技術１における一次ストレージ１に対応する。また、本実施の形態における二次ストレージ１３４は、前提技術１における二次ストレージ２に対応する。また、本実施の形態におけるユーザアプリケーション１１１は、前提技術１におけるアプリケーション部１１に対応する。また、本実施の形態におけるＦＳ制御サーバ１１２は、前提技術１におけるファイルシステム制御部１２に対応する。また、本実施の形態におけるストレージ管理サーバ１３１は、前提技術１における名前空間複製部１３、名前空間追随部１４、マイグレート決定部１６に対応する。また、本実施の形態における名前空間複製ＤＢ１３２は、前提技術１における名前空間複製ＤＢ１５に対応する。

【０１３５】

ＡＣ１２１は、ユーザアプリケーション１１１からの要求を受け付ける。ＭＤＳ１２２は、ノード間排他用トークンのサーバ機能とともに、一次ストレージ１３３のメタデータ（名前空間、ｅｘｔｅｎｔ情報、ｉｎｏｄｅ情報など）を集中管理する。ＨＳＭＡ１２３は、ストレージ管理サーバ１３１からＦＳ制御サーバ１１２への要求を仲介するエージェントプロセスである。ストレージ管理サーバ１３１は、一次ストレージ１３３と二次ストレージ１３４との間のデータコピー機能、両ストレージの空き領域制御などのデバイス制御機能、ファイルシステムとストレージのポリシ制御機能を持つ。

【０１３６】

一次ストレージ１３３は、ファイル１４２、ＤＢ維持フラグ（データベース維持情報）１４３を格納する。ＤＢ維持フラグ１４３は、一次ストレージ１３３のディスクの先頭のスーパーブロックに設定される。二次ストレージは、アーカイブファイル１４４を格納する。また、名前空間複製ＤＢ１３２は、名前空間テーブル１５１とアーカイブＩＤテーブル１５２を格納する。

【０１３７】

次に、名前空間追従処理について、図８のシーケンスを用いて説明する。

【０１３８】

ここで、名前空間複製ＤＢ１３２に対する動作を表す名前空間複製モードとして、通常時に名前空間追従処理を行うイベント通知モードとイベント喪失時などに名前空間複製ＤＢ修正処理を行う修正指令モードがある。名前空間追従処理は、前提技術１と同様であり、ストレージ管理サーバ１３１が名前空間複製ＤＢ１３２の複製を作成した後、ＦＳ制御サーバ１１２からのイベント通知により名前空間複製ＤＢ１３２を更新する処理である。名前空間複製ＤＢ修正処理は、ストレージ管理サーバ１３１がＦＳ制御サーバ１１２に必要な情報を要求することにより名前空間複製ＤＢ１３２の修正を行う処理である。

【０１３９】

まず、ユーザアプリケーション１１１は、名前空間を更新する要求（ｍｋｄｉｒ，ｒｅｎａｍｅ，ｒｍｄｉｒなど）が発生すると、この要求をＦＳ制御サーバ１１２へ送る（Ｓ５１１）。次に、ＡＣ１２１は、受け取った要求をＭＤＳ１２２に送る（Ｓ５１２）。次に、ＭＤＳ１２２は、受け取った要求に従って、一次ストレージ１３３の名前空間を更新し（Ｓ５１３）、一次ストレージ１３３に反映された更新内容をイベント（名前空間更新情報：名前空間遷移イベントやアーカイブ無効化イベント）としてイベントキュー１２４に溜める。所定の時間が経過した後、ＭＤＳ１２２は、イベントキュー１２４に溜まったイベントを事後イベント非同期通知としてストレージ管理サーバ１３１に送る（Ｓ５１４）。次に、ストレージ管理サーバ１３１は、受け取った事後イベント非同期通知に従って名前空間複製ＤＢ１３２の更新を行う（Ｓ５１５）。

【０１４０】

また、所定のポリシや管理者の指示に基づいてアーカイブを行う場合、ストレージ管理サーバ１３１は、ＭＤＳ１２２に滞留しているイベントのフラッシュの要求をＦＳ制御サ

ーバ１１２へ送る（Ｓ５２１）。次に、ＨＳＭＡ１２３は、受け取った要求をＡＣ１２１に送る（Ｓ５２２）。次に、ＡＣ１２１は、受け取った要求をＭＤＳ１２２に送る（Ｓ５２３）。次に、ＭＤＳ１２２は、受け取った要求に従って、イベントキュー１２４に溜まったイベントを事後イベント非同期通知としてストレージ管理サーバ１３１に送る（Ｓ５２４）。

【０１４１】

次に、ストレージ管理サーバ１３１は、受け取った事後イベント非同期通知に従って名前空間複製ＤＢ１３２の更新を行い（Ｓ５２５）、前提技術１のマイグレート決定部１６と同様の処理により、更新された名前空間複製ＤＢ１３２からアーカイブ対象ファイルを検索し（Ｓ５２６）、決定したアーカイブ対象ファイルのアーカイブの要求をＦＳ制御サーバ１１２へ送る（Ｓ５３１）。次に、ＨＳＭＡ１２３は、受け取った要求をＡＣ１２１に送る（Ｓ５３２）。次に、ＡＣ１２１は、受け取った要求をＭＤＳ１２２に送る（Ｓ５３３）。次に、ＭＤＳ１２２は、受け取った要求に従って、メタデータの更新を行い、その結果をストレージ管理サーバ１３１に通知する（Ｓ５３４）。次に、ストレージ管理サーバ１３１は、二次ストレージ１３４にアーカイブを作成する（Ｓ５３５）。

【０１４２】

次に、名前空間複製モードを決定する名前空間複製モード決定処理について説明する。

【０１４３】

図９は、本実施の形態に係る名前空間複製モード決定処理の動作の一例を示すフローチャートである。このフローチャートの左半分は修正指令モードの動作を示し、右半分はイベント通知モードの動作を示す。また、ＤＢ維持フラグ１４３は、セットされていれば名前空間複製ＤＢ修正処理が不要であることを表す。

【０１４４】

まず、ＭＤＳ１２２は、整然起動を行う、またはフェールオーバによる起動を行う（Ｓ６１１）。このとき、名前空間複製モードは修正指令モードである。次に、ＭＤＳ１２２は、一次ストレージ１３３内のＤＢ維持フラグ１４３がセットされているか否かの判断を行う（Ｓ６１２）。

【０１４５】

ＤＢ維持フラグ１４３がセットされている場合（Ｓ６１２，Ｙ）、ＭＤＳ１２２は、名前空間複製モードを修正指令モードからイベント通知モードに変更し、一旦ＤＢ維持フラグ１４３をクリアし（Ｓ６２２）、通常処理を行う。ＭＤＳ１２２は、通常処理の過程でイベント消失が検出されず（Ｓ６２３，Ｎ）、停止要求もなければ、通常処理を続行する。ここで、ＭＤＳ１２２は、停止要求があれば（Ｓ６２４、Ｙ）。停止処理を行う。この停止処理を整然と完遂できると判断した場合（Ｓ６２５，Ｙ）、その処理過程で、ＤＢ維持フラグ１４３をセットし（Ｓ６２６）、このフローを終了する。このＤＢ維持フラグ１４３のセットにより、次のＭＤＳ１２２の起動が整然起動であると認識される。

【０１４６】

通常処理中にイベント消失が検出された場合（Ｓ６２３，Ｙ）、ＭＤＳ１２２は、名前空間複製モードをイベント通知モードから修正指令モードに変更し、修正指令をストレージ管理サーバ１３１に送信することにより、ストレージ管理サーバ１３１に名前空間複製ＤＢ修正処理を実行させる。

【０１４７】

処理Ｓ６１２において、ＤＢ維持フラグ１４３がクリアされている場合（Ｓ６１２，Ｎ）、ＭＤＳ１２２は、修正指令をストレージ管理サーバ１３１に送信することにより、ストレージ管理サーバ１３１に名前空間複製ＤＢ修正処理を実行させる。ＭＤＳ１２２は、修正指令に対する応答の待機中に（Ｓ６１３、Ｙ）、停止要求もなければ（Ｓ６１４、Ｎ）、引き続き応答の待機を行う。また、ＭＤＳ１２２は、応答の待機中に（Ｓ６１３、Ｙ）、停止要求があれば（Ｓ６１４、Ｙ）。停止処理を行い、このフローを終了する。また、ＭＤＳ１２２は、修正指令に対する正常応答を受信した場合（Ｓ６１３，Ｙ）、名前空間複製モードを修正指令モードからイベント通知モードに変更し、通常処理を行う。

【０１４８】
次に、名前空間に関するデータ構造について説明する。
【０１４９】
図１０は、本実施の形態に係る名前空間に関するデータ構造の一例を示すブロック図である。この図は、一次ストレージ１３３、二次ストレージ１３４、名前空間複製ＤＢ１３２のデータ構造を表す。
【０１５０】
一次ストレージ１３３において、各ファイルは、ｉｎｏｄｅ情報（図中では一次ストレージ１３３内の丸で表される）とファイルデータ（図中では一次ストレージ１３３内の四角で表される）からなる。ｉｎｏｄｅ情報は、ｉｎｏｄｅ番号、ｇｅｎ番号、属性、時刻情報からなる。ｇｅｎ（generation）番号は、同一ｉｎｏｄｅ番号を持つファイルを世代で識別するための番号であり、ＮＦＳ（Network File System）やＨＳＭで用いられるものである。属性は、ファイルの種類がディレクトリファイルであるか通常ファイルであるか、などの情報である。時刻情報は、ｍｔｉｍｅ（データ更新時刻）、ｃｔｉｍｅ（ｉｎｏｄｅ更新時刻）、ａｔｉｍｅ（アクセス時刻）からなる。ｉｎｏｄｅ情報を更新するとｃｔｉｍｅも更新される。
【０１５１】
また、一次ストレージ１３３におけるファイルには、ディレクトリファイルと通常ファイルがある。一次ストレージ１３３におけるディレクトリファイルのファイルデータは、子のファイルへのリンク毎のリンク情報を持つ。リンク情報は、１つの子のファイルの名前とｉｎｏｄｅ番号からなる。また、一次ストレージ１３３における通常ファイルのファイルデータは、通常ファイルデータ、またはアーカイブＩＤである。
【０１５２】
ｉｎｏｄｅ番号＝８のディレクトリファイルの下に、ｉｎｏｄｅ番号＝９のディレクトリファイルとｉｎｏｄｅ番号＝１０のディレクトリファイルが存在する。ｉｎｏｄｅ番号＝９のディレクトリファイルの下に、ｉｎｏｄｅ番号＝１１の通常ファイルが存在する。ｉｎｏｄｅ番号＝１０のディレクトリファイルの下に、ｉｎｏｄｅ番号＝１２の通常ファイルが存在する。
【０１５３】
ｉｎｏｄｅ番号＝８，９，１０のディレクトリファイルは、親のｉｎｏｄｅ番号、子の名前、子のｉｎｏｄｅ番号を持つ。ｉｎｏｄｅ番号＝１１の通常ファイルは、ｉｎｏｄｅ番号＝９とｉｎｏｄｅ番号＝１０の両方のディレクトリからリンクされており、二次ストレージ１３４にアーカイブされているためファイルデータとしてアーカイブＩＤを持つ。ｉｎｏｄｅ番号＝１２の通常ファイルは、ファイルデータとして通常ファイルデータを持つ。
【０１５４】
名前空間複製ＤＢ１３２の名前空間テーブル１５１は、一次ストレージ１３３の名前空間をデータベースで表したものであり、一次ストレージ１３３におけるリンク毎のエントリが作成され、親のディレクトリファイル毎にまとめて保存される。親がディレクトリファイルで子がディレクトリファイルであるリンクのエントリ（親がｉｎｏｄｅ番号＝８で子がｉｎｏｄｅ番号＝９、親がｉｎｏｄｅ番号＝８で子がｉｎｏｄｅ番号＝１０）は、親のｉｎｏｄｅ番号（ｇｅｎ番号）、子の名前、子のｉｎｏｄｅ番号（ｇｅｎ番号）を持つ。また、親がディレクトリファイルで子が通常ファイルであるリンクのエントリは（親がｉｎｏｄｅ番号＝９で子がｉｎｏｄｅ番号＝１１、親がｉｎｏｄｅ番号＝１０で子がｉｎｏｄｅ番号＝１２、親がｉｎｏｄｅ番号＝１０で子がｉｎｏｄｅ番号＝１１）、親のｉｎｏｄｅ番号（ｇｅｎ番号）、ポリシＩＤ、子の名前、子のｉｎｏｄｅ番号、子の最終アクセス時刻、ポリシ制御における子の状態値、子のアーカイブＩＤなど、子のファイルについての詳細な情報を含む。
【０１５５】
名前空間複製ＤＢ１３２のアーカイブＩＤテーブル１５２は、二次ストレージ１３４上

の論理的な位置に対応するアーカイブＩＤを管理するものであり、ファイル毎のエントリが作成される。エントリは、アーカイブＩＤ、アーカイブデータ状態値、リコールＩＤ、最終データ更新時刻、リストア時のｉｎｏｄｅ情報作成用情報を持つ。

【０１５６】

二次ストレージ１３４は、アーカイブファイル毎に、アーカイブＩＤ、パス情報、属性情報、ファイルデータを持つ。この例におけるパス情報は、ポリシＢでアーカイブされた場合のパス情報である。

【０１５７】

次に、イベントの種類と内容について説明する。

【０１５８】

図１１は、本実施の形態に係るイベントの種類と内容の一例を示す表である。ＦＳ制御サーバ１１２からストレージ管理サーバ１３１へ通知されるイベントの種類には、名前挿入、名前除去、名前変更、ｉｎｏｄｅ情報変更がある。

【０１５９】

名前挿入は、ディレクトリへの名前挿入を伴うメタデータ処理、つまり、親のディレクトリファイルに子のファイルの名前やｉｎｏｄｅ番号などのリンク情報を挿入することを示す。名前除去は、ディレクトリからの名前削除を伴うメタデータ処理、つまり、親のディレクトリファイルからリンク情報を削除することを示す。名前変更は、ディレクトリを跨ぐ、跨がないに関わらず、名前変更を伴うメタデータ処理、つまり、リンク情報を移動することを示す。ｉｎｏｄｅ情報変更は、リンク情報の変更がなく、ｉｎｏｄｅ情報の変更を伴うメタデータ処理、つまり、あるファイルに書き込みが発生してｉｎｏｄｅ情報のｍｔｉｍｅが変更されたことなどを示す。

【０１６０】

次に、イベントの内容（名前空間変更内容）として付加されるイベント付加情報の種類には、親ｉｎｏｄｅ番号（その１）、親ｉｎｏｄｅ番号（その２）、対象ファイル名（その１）、対象ファイル名（その２）、子ｉｎｏｄｅ情報、イベント発生時刻（名前空間変更時刻）がある。これらのイベント付加情報のうち、親ｉｎｏｄｅ番号（その１）、親ｉｎｏｄｅ番号（その２）、子ｉｎｏｄｅ情報は、ｉｎｏｄｅ／ｇｅｎ番号、ｃｔｉｍｅ／ｍｔｉｍｅ／ａｔｉｍｅ、ｅｘｔｅｎｔ情報などを持つ。

【０１６１】

この表は、イベントの種類毎の列とイベント付加情報毎の行とで表され、あるイベントの種類の内容にあるイベント付加情報が含まれる場合、交差する欄に“○”が付けている。名前挿入と名前除去は、親ｉｎｏｄｅ番号（その１）、対象ファイル名（その１）、子ｉｎｏｄｅ情報、イベント発生時刻を含む。名前変更は、親ｉｎｏｄｅ番号（その１）、親ｉｎｏｄｅ番号（その２）、対象ファイル名（その１）、対象ファイル名（その２）、子ｉｎｏｄｅ情報、イベント発生時刻を含む。ｉｎｏｄｅ情報変更は、子ｉｎｏｄｅ情報、イベント発生時刻を含む。

【０１６２】

名前変更におけるイベント付加情報のみ、親ｉｎｏｄｅ番号と対象ファイル名が２つずつ存在するが、親ｉｎｏｄｅ番号（その１）、対象ファイル名（その１）が変更前を示し、親ｉｎｏｄｅ番号（その２）、対象ファイル名（その２）が変更後を示す。また、子ｉｎｏｄｅ情報とイベント発生時刻は、全てのイベントに付加される。

【０１６３】

その他、イベントと同様、ＦＳ制御サーバ１１２からストレージ管理サーバ１３１へ通知されるものとして修正指令がある。修正指令は、上述した名前空間複製モード決定処理により発生する。

【０１６４】

次に、名前空間複製ＤＢ修正処理について説明する。

【０１６５】

一次ストレージ１３３のファイルシステムにおいて、ｉｎｏｄｅ情報が更新されると、

必ずそのｉｎｏｄｅ情報のｃｔｉｍｅも更新される。リンク情報が更新されると、そのリンクの両端のｉｎｏｄｅ情報も更新され、それらのｉｎｏｄｅ情報のｃｔｉｍｅも更新される。一方、上述したように、通常処理において、ＦＳ制御サーバ１１２は、イベントの内容とともにイベント発生時刻をストレージ管理サーバ１３１に通知する。ストレージ管理サーバ１３１は、名前空間複製ＤＢ１３２に反映された最終のイベント発生時刻を最終イベント発生時刻として記憶する。

【０１６６】

従って、ストレージ管理サーバ１３１は、最終イベント発生時刻より後のｃｔｉｍｅを持つｉｎｏｄｅ情報とそのｉｎｏｄｅ情報から子へのリンク情報を用いて名前空間複製ＤＢ１３２を修正すれば良い。ここでは、わずかな時刻のずれを考慮し、ストレージ管理サーバ１３１は、「最終イベント発生時刻より後の」ではなく「最終イベント発生時刻以上の」ｃｔｉｍｅを持つｉｎｏｄｅ情報を用いる。

【０１６７】

以下、名前空間複製ＤＢ修正処理について具体例を用いて説明する。図１２は、本実施の形態に係る名前空間ＤＢ修正処理の動作の一例を示すシーケンス図である。この図は、ＦＳ制御サーバ１１２とストレージ管理サーバ１３１の動作を示す。また、図１３は、イベント喪失が発生した時点における一次ストレージの名前空間の内容の一例を示すツリー構造図である。各ノードは、ファイル毎のｉｎｏｄｅ情報を表す。このうち、丸で表されたノードは、ディレクトリファイルのｉｎｏｄｅ情報を表し、四角で表されたノードは通常ファイルのｉｎｏｄｅ情報を表す。また、ノードに記された数字は、ｉｎｏｄｅ情報におけるｃｔｉｍｅの値を表す。また、ノード間を結ぶ線はリンク情報を表す。

【０１６８】

まず、通常処理として、ＦＳ制御サーバ１１２は、時刻ｔ＝１０のイベント通知（Ｓ７１１）、時刻ｔ＝２０のイベント通知（Ｓ７１２），時刻ｔ＝３０のイベント通知（Ｓ７１３）を行う。この具体例において、ｔ＝１０においては、ｃｔｉｍｅ＝１０，１０のイベントが通知され、ｔ＝２０においては、ｃｔｉｍｅ＝１５，１５，１５，２０，２０のイベントが通知される。また、ｔ＝３０のイベント通知は、通信障害によりストレージ管理サーバ１３１に届かなかったとする。図１４は、イベント喪失が発生した時点における名前空間テーブルの内容の一例を示すツリー構造図である。一次ストレージ１３３の名前空間と比較すると、名前空間テーブルは、ｃｔｉｍｅ＝２０より後のｉｎｏｄｅ情報やリンク情報が欠落している。

【０１６９】

その後、ストレージ管理サーバ１３１は、ＦＳ制御サーバ１１２の名前空間複製モード決定処理により修正指令を通知されると（Ｓ７２０）、名前空間複製ＤＢ１３２におけるｉｎｏｄｅ情報の修正を行うｉｎｏｄｅ情報修正処理として、ｃｔｉｍｅが２０以上であるｉｎｏｄｅ情報（未更新ｉｎｏｄｅ情報）の要求（未更新ｉｎｏｄｅ情報要求）をＦＳ制御サーバ１１２へ送る（Ｓ７２１）。ＦＳ制御サーバ１１２は、この要求に従ってｃｔｉｍｅが２０以上のｉｎｏｄｅ情報を列挙し、列挙したｉｎｏｄｅ情報を対象ｉｎｏｄｅ情報としてストレージ管理サーバ１３１へ送る（Ｓ７２２）。この具体例においては、ｃｔｉｍｅ＝２０，２５，３５，３５のｉｎｏｄｅ情報がストレージ管理サーバ１３１へ送られる。

【０１７０】

ストレージ管理サーバ１３１は、受け取った対象ｉｎｏｄｅ情報を用いて名前空間テーブル１５１の修正を行う。図１５は、ｉｎｏｄｅ情報が修正された時点における名前空間テーブルの内容の一例を示すツリー構造図である。太枠で囲われたノードは、修正されたｉｎｏｄｅ情報に対応する。その他のノードは、現状維持のｉｎｏｄｅ情報に対応する。この時点で、名前空間テーブル１５１は、リンク情報が張られていないｉｎｏｄｅ情報の存在を許している。

【０１７１】

ここで、ＦＳ制御サーバ１１２からストレージ管理サーバ１３１へ、新たなイベントが

通知されると（Ｓ７３０）、ストレージ管理サーバ１３１は、通知されたイベントが既に修正されたｉｎｏｄｅ情報に関係するか否かの判断を行い、無関係であれば、随時、そのイベントを名前空間テーブル１５１に反映する。図１６は、修正されたｉｎｏｄｅ情報と無関係のイベントが反映された時点における名前空間テーブルの内容の一例を示すツリー構造図である。太枠で囲われたノードは、修正されたｉｎｏｄｅ情報と無関係のイベントが反映されたｉｎｏｄｅ情報に対応する。

【０１７２】

次に、ストレージ管理サーバ１３１は、名前空間複製ＤＢ１３２におけるリンク情報の修正を行うリンク情報修正処理として、修正されたｉｎｏｄｅ情報のうち、ディレクトリファイルのｉｎｏｄｅ情報を抽出し、抽出されたディレクトリファイルが持つリンク情報（未更新リンク情報）の要求（未更新リンク情報要求）をＦＳ制御サーバ１１２へ送る（Ｓ７３１）。ＦＳ制御サーバ１１２は、この要求に従って未更新リンク情報を列挙し、列挙した未更新リンク情報をストレージ管理サーバ１３１へ送る（Ｓ７３２）。このときＦＳ制御サーバ１１２は、未更新リンク情報と共に、未更新リンク情報に示された子のファイルのｉｎｏｄｅ情報も送る。この具体例においては、ディレクトリファイルであるｃｔｉｍｅ＝２０，３５のｉｎｏｄｅ情報に対応するリンク情報がストレージ管理サーバ１３１へ送られる。

【０１７３】

ストレージ管理サーバ１３１は、受け取った未更新リンク情報に基づいて、名前空間テーブル１５１におけるリンク情報の修正を行い、名前空間複製ＤＢ修正処理が終了する。図１７は、リンク情報が修正された時点における名前空間テーブルの内容の一例を示すツリー構造図である。太枠で囲われたノードは、修正されたｉｎｏｄｅ情報のうちディレクトリファイルであるｉｎｏｄｅ情報に対応し、太線で表されたリンク情報は、修正されたリンク情報に対応する。

【０１７４】

なお、処理Ｓ７３０において、新たなイベントが修正されたｉｎｏｄｅ情報に無関係である場合について述べたが、新たなイベントが修正されたｉｎｏｄｅ情報に関係する場合、関係のあるｉｎｏｄｅ情報のｃｔｉｍｅと新たなイベントに含まれたイベント発生時刻との比較を行い、新しい方の情報を用いてｉｎｏｄｅ情報やリンク情報の修正を行う。

【０１７５】

上述したように、ストレージ管理サーバ１３１は、リンク情報毎のエントリを持つデータベースである名前空間テーブル１５１として名前空間を複製する。この名前空間テーブル１５１によれば、常にツリー構造が完全な状態でならなくてはならない通常の名前空間とは異なり、ツリー構造が不完全な状態からの修正が容易になる。

【０１７６】

前提技術１においては、イベント喪失時に名前空間のツリーを全てスキャンして名前空間複製ＤＢを作成し直し、新たなイベントの発生の度にイベントに関連するツリーをスキャンして名前空間複製ＤＢを修正していた。一方、本実施の形態によれば、イベント喪失時、ストレージ管理サーバ１３１は、名前空間複製ＤＢ１３２に反映されていない未更新ｉｎｏｄｅ情報と未更新リンク情報だけを用いて名前空間複製ＤＢ１３２の修正を行うため、負荷が小さく高速に名前空間複製ＤＢ１３２の修正を行うことができる。また、ストレージ管理サーバ１３１は、名前空間複製ＤＢの修正中に新たなイベントが通知されても、そのイベントと未更新ｉｎｏｄｅ情報のうち新しい方を用いて名前空間複製ＤＢに反映することにより、負荷が小さく高速に名前空間複製ＤＢ１３２の更新を行うことができる。従って、巨大なファイルシステムにおける名前空間の複製が可能になる。

【０１７７】

また、本実施の形態に係る名前空間複製装置は、ストレージシステムに容易に適用することができ、ストレージシステムの性能をより高めることができる。ここで、ストレージシステムには、例えばＨＳＭシステム、バックアップシステム等が含まれ得る。

【０１７８】

更に、名前空間複製装置を構成するコンピュータにおいて上述した各ステップを実行させるプログラムを、名前空間複製プログラムとして提供することができる。上述したプログラムは、コンピュータにより読取り可能な記録媒体に記憶させることによって、名前空間複製装置を構成するコンピュータに実行させることが可能となる。ここで、上記コンピュータにより読取り可能な記録媒体としては、ＲＯＭやＲＡＭ等のコンピュータに内部実装される内部記憶装置、ＣＤ－ＲＯＭやフレキシブルディスク、ＤＶＤディスク、光磁気ディスク、ＩＣカード等の可搬型記憶媒体や、コンピュータプログラムを保持するデータベース、或いは、他のコンピュータ並びにそのデータベースや、更に回線上の伝送媒体をも含むものである。

【０１７９】

なお、ストレージ装置は、実施の形態における一次ストレージに対応する。また、ファイルシステム制御装置は、実施の形態におけるＦＳ制御サーバに対応する。

【０１８０】

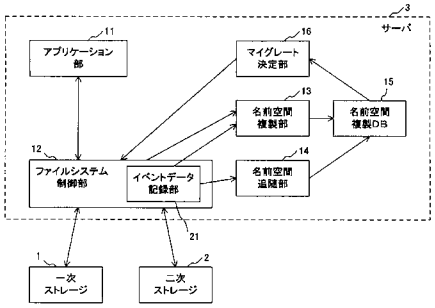
また、名前空間複製データベース更新ステップは、実施の形態における名前空間追従処理に対応する。また、名前空間複製データベース修正ステップは、実施の形態における名前空間複製データベース修正処理に対応する。また、名前空間複製データベース更新部は、実施の形態におけるストレージ管理サーバにおける名前空間追従処理に対応する。また、名前空間複製データベース修正部は、実施の形態におけるストレージ管理サーバにおける名前空間複製データベース修正処理に対応する。

【産業上の利用可能性】

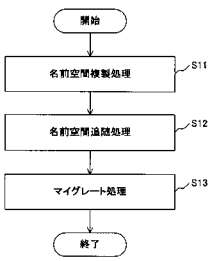
【０１８１】

以上説明したように、本発明によれば、ストレージ装置上の名前空間をデータベースとして効率的に複製することができる。

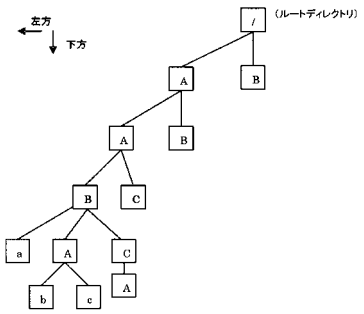
【図１】



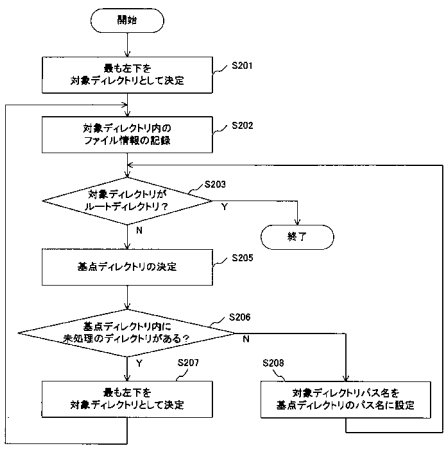
【図２】



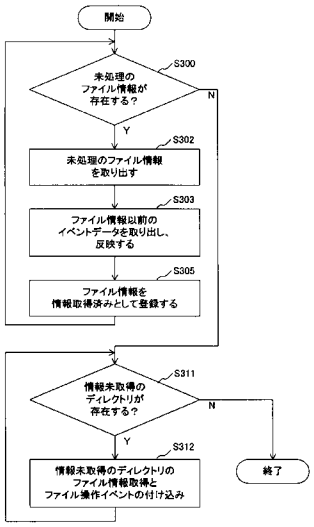
【図３】



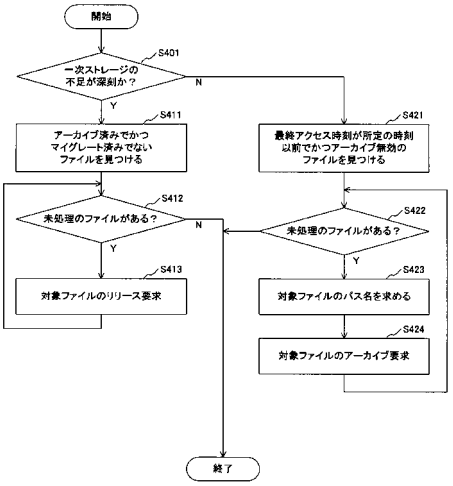
【図４】



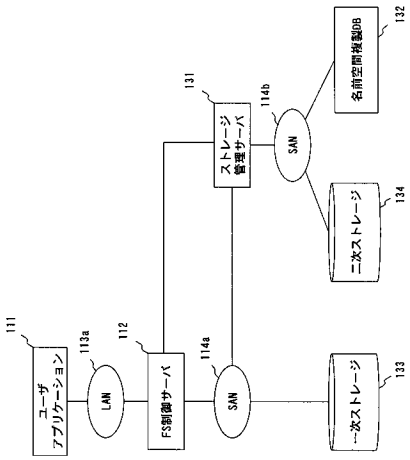
【図５】


開始
S300
未処理の
ファイル情報が
存在する？
N
Y
S302
未処理のファイル情報
を取り出す
S303
ファイル情報以前の
イベントデータを取り出し、
反映する
S305
ファイル情報を
情報取得済みとして登録する
S311
情報未取得の
ディレクトリが
存在する？
N
Y
S312
情報未取得のディレクトリの
ファイル情報取得と
ファイル操作イベントの付け込み
終了

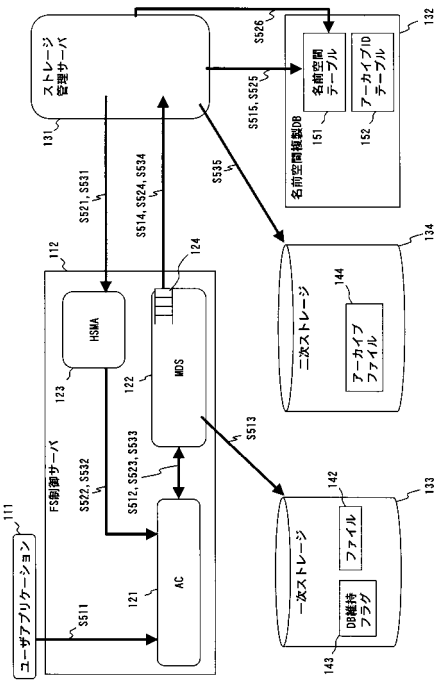
【図６】


開始
S401
一次ストレージの
不足が深刻か？
N
Y
S411
アーカイブ済みでかつ
マイグレート済みでない
ファイルを見つける
S412
未処理のファイルがある？
N
Y
S413
対象ファイルのリリース要求
S421
最終アクセス時刻が所定の時刻
以前でかつアーカイブ無効の
ファイルを見つける
S422
未処理のファイルがある？
Y
S423
対象ファイルのパス名を求める
S424
対象ファイルのアーカイブ要求
終了

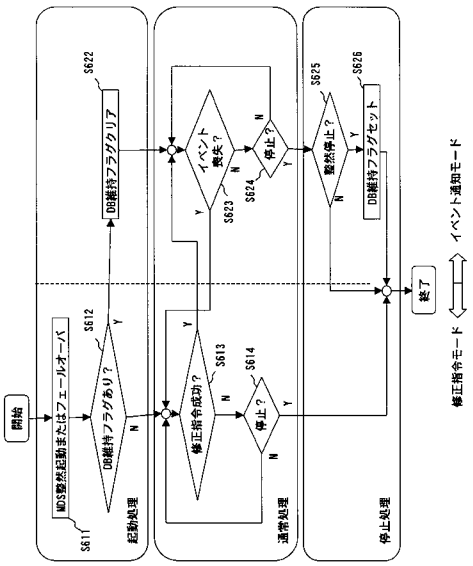
【図７】


111
ユーザ
アプリケーション
113a
LAN
112
FS制御サーバ
114a
SAN
131
ストレージ
管理サーバ
114b
SAN
132
名前空間複製DB
134
二次ストレージ
133
一次ストレージ

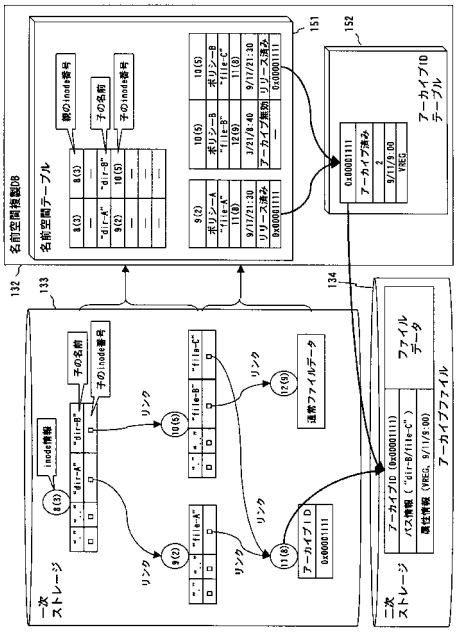
【図８】


ストレージ
管理サーバ
131
S526
S515, S525
名前空間
テーブル
アーカイブID
テーブル
名前空間複製DB
151
152
132
S535
S521, S531
S514, S524, S534
112
124
HSMA
123
MDS
122
二次ストレージ
144
アーカイブ
ファイル
134
S513
FS制御サーバ
S522, S532
S512, S523, S533
AC
121
一次ストレージ
142
ファイル
DB維持
フラグ
143
133
111
ユーザアプリケーション
S511

【図９】



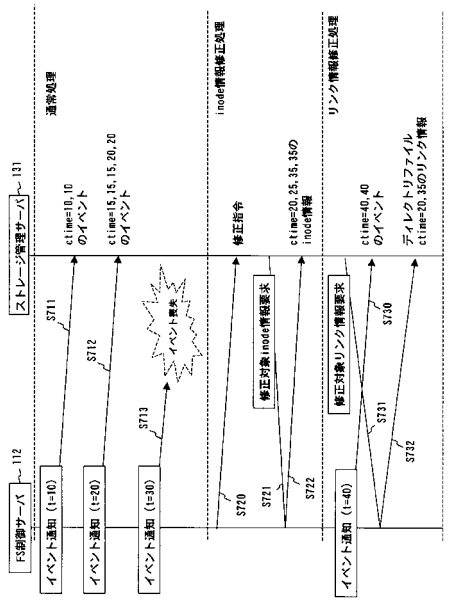
【図１０】



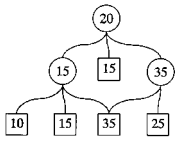
【図１１】

	名前挿入	名前除去	名前変更	inode情報変更
親inode情報（その1）	○	○	○	
親inode情報（その2）			○	
対象ファイル名（その1）	○	○	○	
対象ファイル名（その2）			○	
子inode情報	○	○	○	○
イベント発生時刻	○	○	○	○

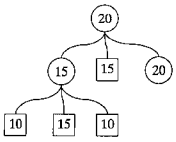
【図１２】



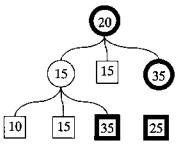
【図１３】



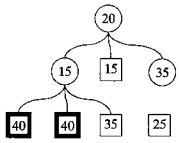
【図１４】



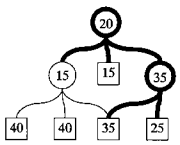
【図１５】



【図１６】



【図１７】



フロントページの続き

(56)参考文献 特開平６－５９９５７（ＪＰ，Ａ）
特開２００５－１９６７２５（ＪＰ，Ａ）
特開２００６－３９８１４（ＪＰ，Ａ）
国際公開第９５／２９４４４（WO，Ａ１）
特開２００３－２８０９５０（ＪＰ，Ａ）
特開平４－４８３６５（ＪＰ，Ａ）
特開２０００－２７６３９１（ＪＰ，Ａ）
特開２０００－２８４９９５（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
G06F 12/00