

(19) 世界知的所有権機関  
国際事務局



(43) 国際公開日  
2007年10月4日 (04.10.2007)

PCT

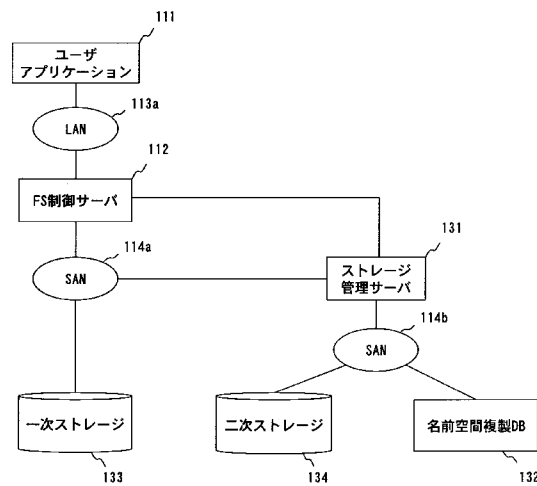
(10) 国際公開番号  
WO 2007/110931 A1

- (51) 国際特許分類: *G06F 12/00* (2006.01)
- (21) 国際出願番号: PCT/JP2006/306284
- (22) 国際出願日: 2006年3月28日 (28.03.2006)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人 (米国を除く全ての指定国について): 富士通株式会社 (FUJITSU LIMITED) [JP/JP]; 〒2118588 神奈川県川崎市中原区上小田中4丁目1番1号 Kanagawa (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 塩沢 賢輔 (SHIOZAWA, Kensuke) [JP/JP]; 〒2118588 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 Kanagawa (JP). 新開 慶武 (SHINKAI, Yoshitake)
- (74) 代理人: 赤澤 日出夫 (AKAZAWA, Hideo); 〒1640001 東京都中野区中野4-5-1 中野K Iビル4階 Tokyo (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD,

[ 続葉有 ]

(54) Title: NAME SPACE COPYING PROGRAM, NAME SPACE COPYING DEVICE, AND NAME SPACE COPYING METHOD

(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



111 USER APPLICATION	133 PRIMARY STORAGE
112 FS CONTROL SERVER	134 SECONDARY STORAGE
131 STORAGE MANAGING SERVER	132 NAME SPACE COPYING DB

(57) Abstract: A name space copying program causes a computer to execute: a name space copying database update step for acquiring an event concerning a name space update from an FS control server (112) controlling a primary storage (133) and updating a name space copying DB (132) created according to inode information and link information in the primary storage (133); and a name space copying database correction step for acquiring inode information having a time of a predetermined time and after link information corresponding to the inode information from the FS control server (112) and correcting the name space copying DB (132) when an event not reflecting the name space copying DB (132) is lost.

(57) 要約: 一次ストレージ133の制御を行うFS制御サーバ112から名前空間更新に関するイベントを取得し、一次ストレージ133におけるinode情報と

[ 続葉有 ]



WO 2007/110931 A1



SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:

— 国際調査報告書

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

---

リンク情報に基づいて作成された名前空間複製DB132を、イベント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、名前空間複製DB132への反映が行われていないイベントが喪失された場合、所定時刻以降のctimeを持つinode情報とそのinode情報に対応するリンク情報とを、FS制御サーバ112から取得し、名前空間複製DB132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。

## 明 細 書

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

### 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

大きくなってしまふ。

- [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ふ。このため、HSM制御装置は、顧客が設定したポリシー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ーヘッドが大幅に増加する。
- [0008] 第3にHSMポリシー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ポリシー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ー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、名前空間追従部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や名前空間追従部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] 名前空間追従部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、名前空間追従部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、ポリシー制御の一例として、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ーに従い、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)、名前空間追隨処理(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の追隨のオーバー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、以下の条件を全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、名前空間追従処理について説明する。

[0098] 名前空間追従部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、名前空間追従部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、名前空間追従部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、名前空間追従部14は削除する。また、名前空間追従部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、名前空間追従部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、名前空間追従部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、ユーザポリシー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従するため、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、名前空間追従部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従部14に通知し、名前空間追従部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従部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に基づくポリシー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ー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から名前空間追従部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、名前空間追従部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ー制御機能を持つ。

- [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、所定のポリシー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ーによる起動を行う(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14

3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(generator)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、ポリシー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“○”が付いている。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。

## 請求の範囲

- [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。
- [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。
- [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。
- [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名前空間複製プログラム。

[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名前空間複製プログラム。

[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。

[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。

[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名前空間複製プログラム。

[9] 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、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名前空間複製プログラム。

[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。

[11] 請求項10に記載の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名前

空間複製装置。

- [12] 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。
- [13] 請求項12に記載の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名前空間複製方法。
- [14] 請求項12に記載の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名前空間複製方法。
- [15] 請求項12に記載の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名前空間複製方法。

[16] 請求項12に記載の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名前空間複製方法。

[17] 請求項12に記載の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名前空間複製方法。

[18] 請求項12に記載の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名前空間複製方法。

[19] 請求項13に記載の名前空間複製方法において、

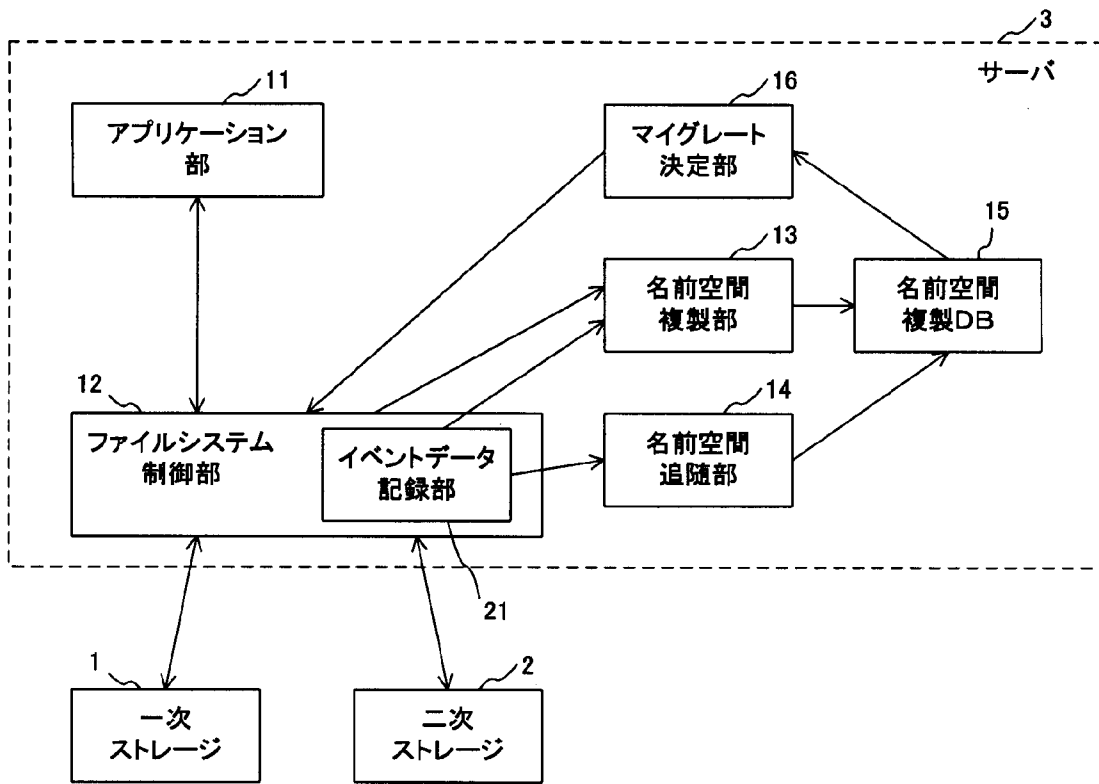
前記名前空間複製データベース情報修正ステップは、前記名前空間複製装置において、前記名前空間複製データベースの修正が完了する前に、前記名前空間複製データベース更新ステップにより前記名前空間複製データベースに反映されていない名前空間更新情報を取得し、該名前空間更新情報が前記未更新ファイル識別

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方法。

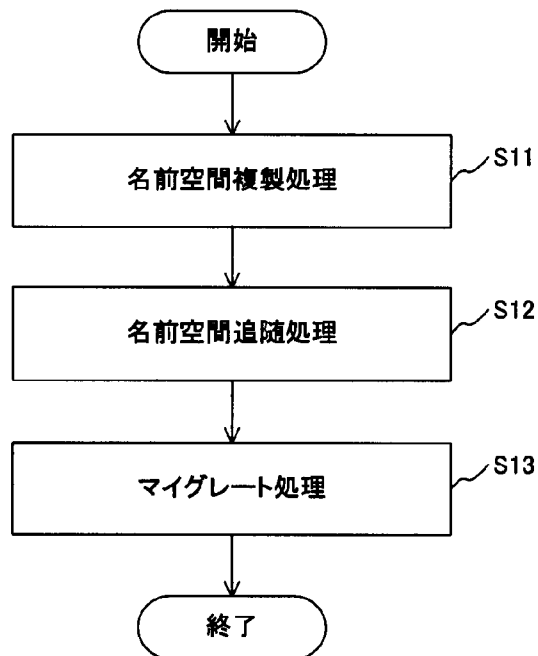
[20] 請求項12に記載の名前空間複製方法において、

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名前空間複製方法。

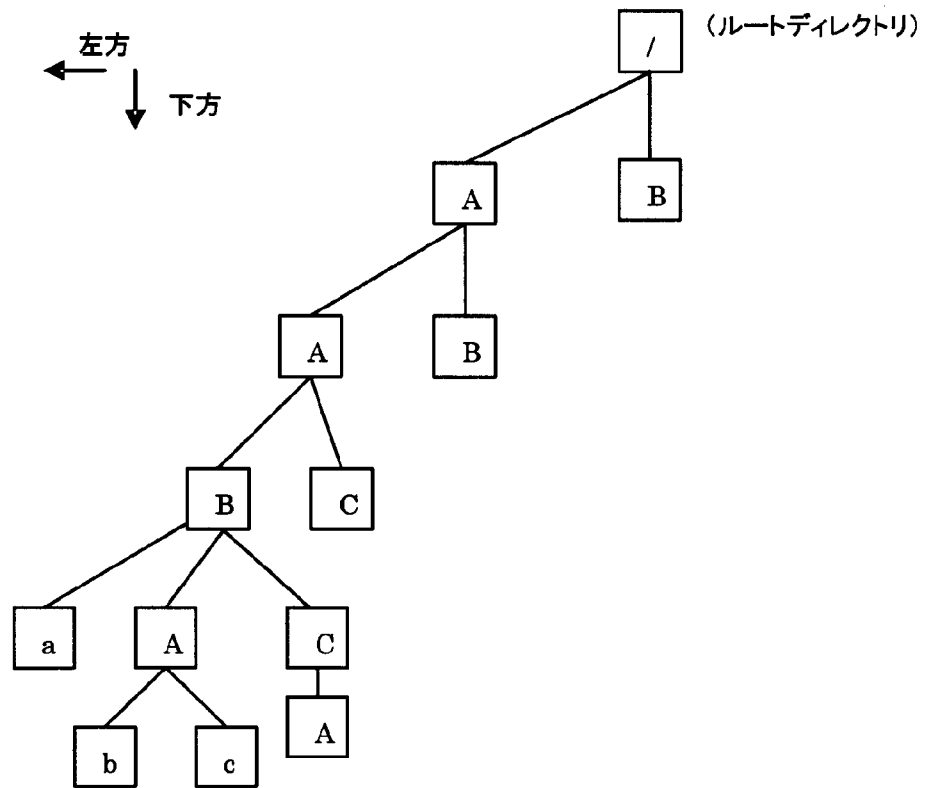
[図1]



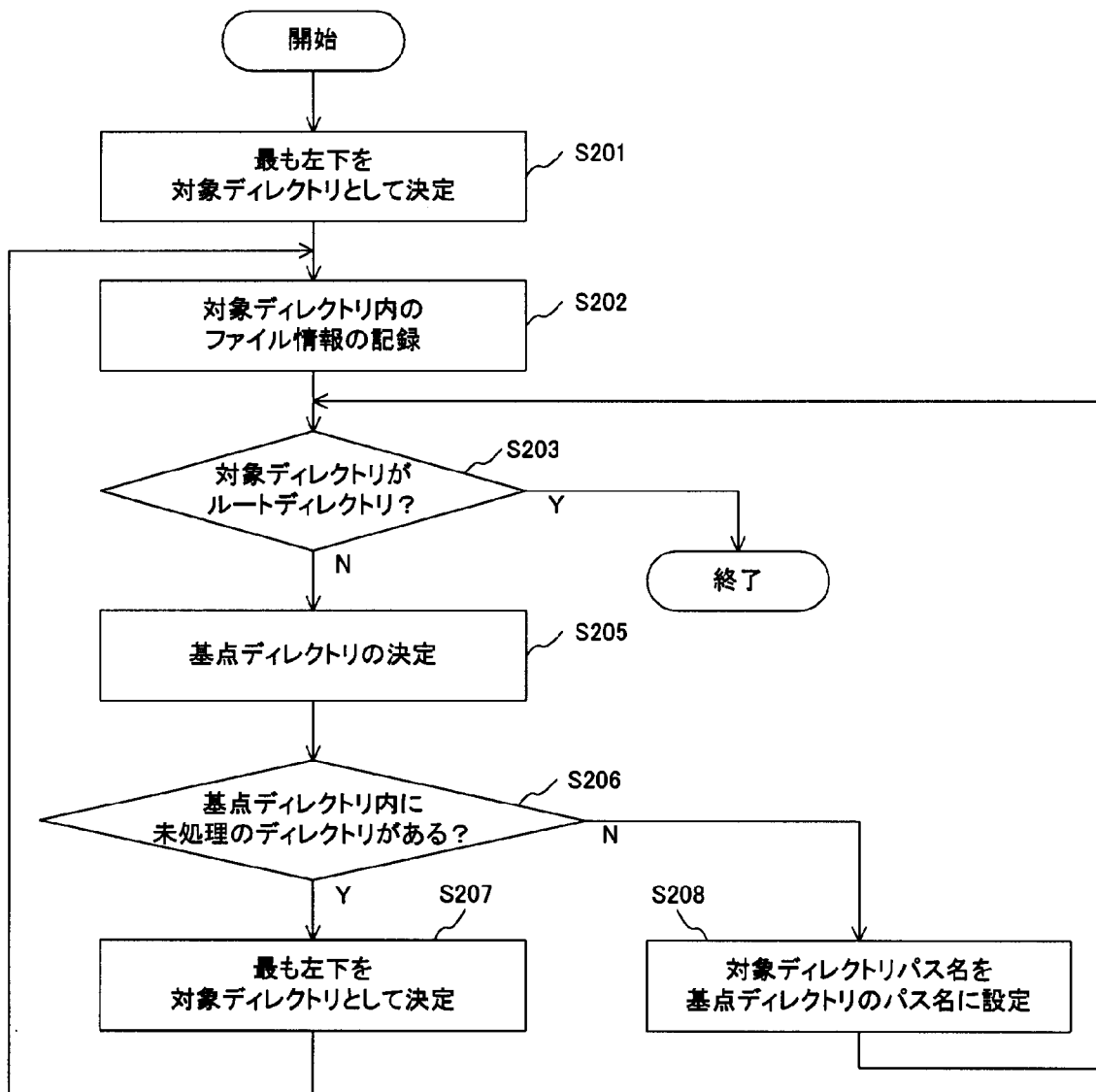
[図2]



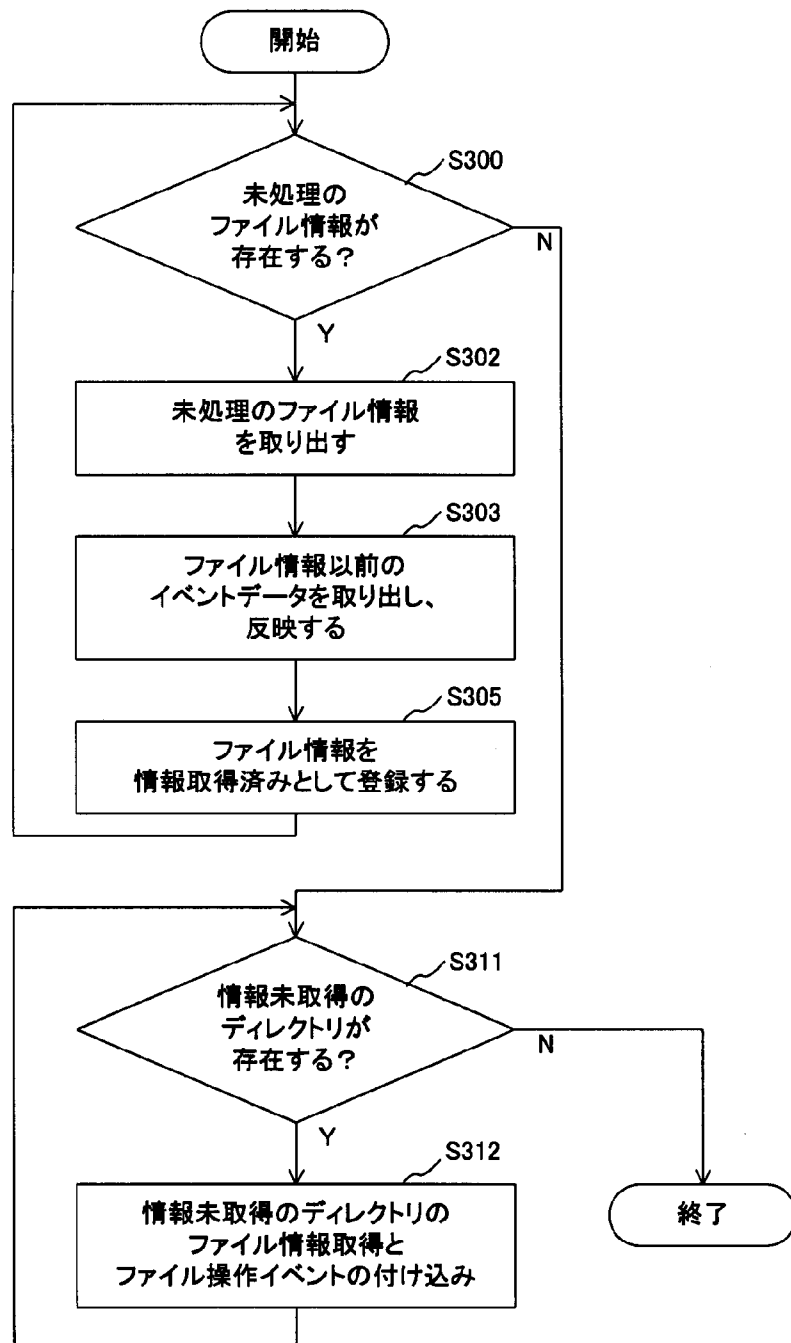
[図3]



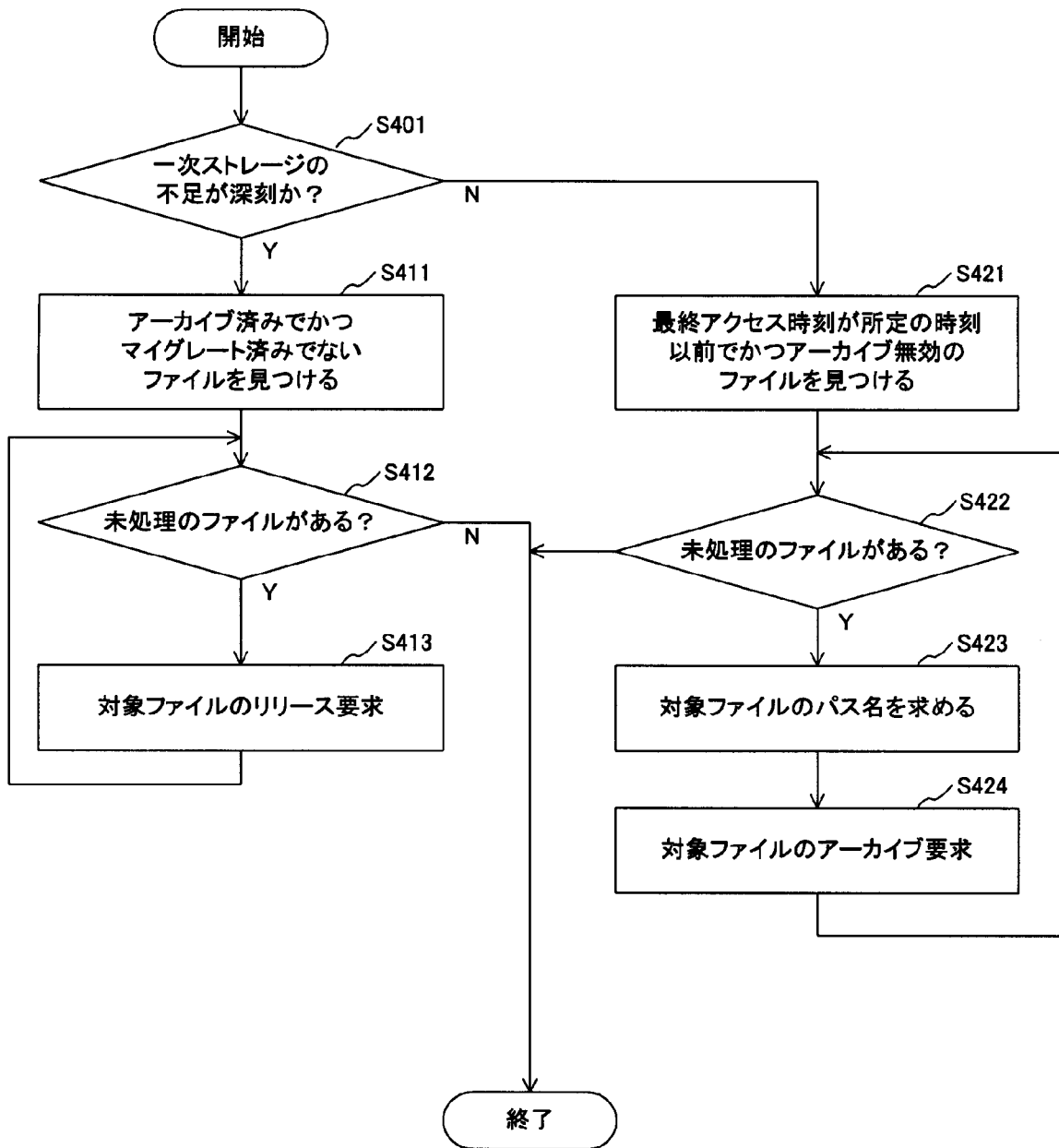
[図4]



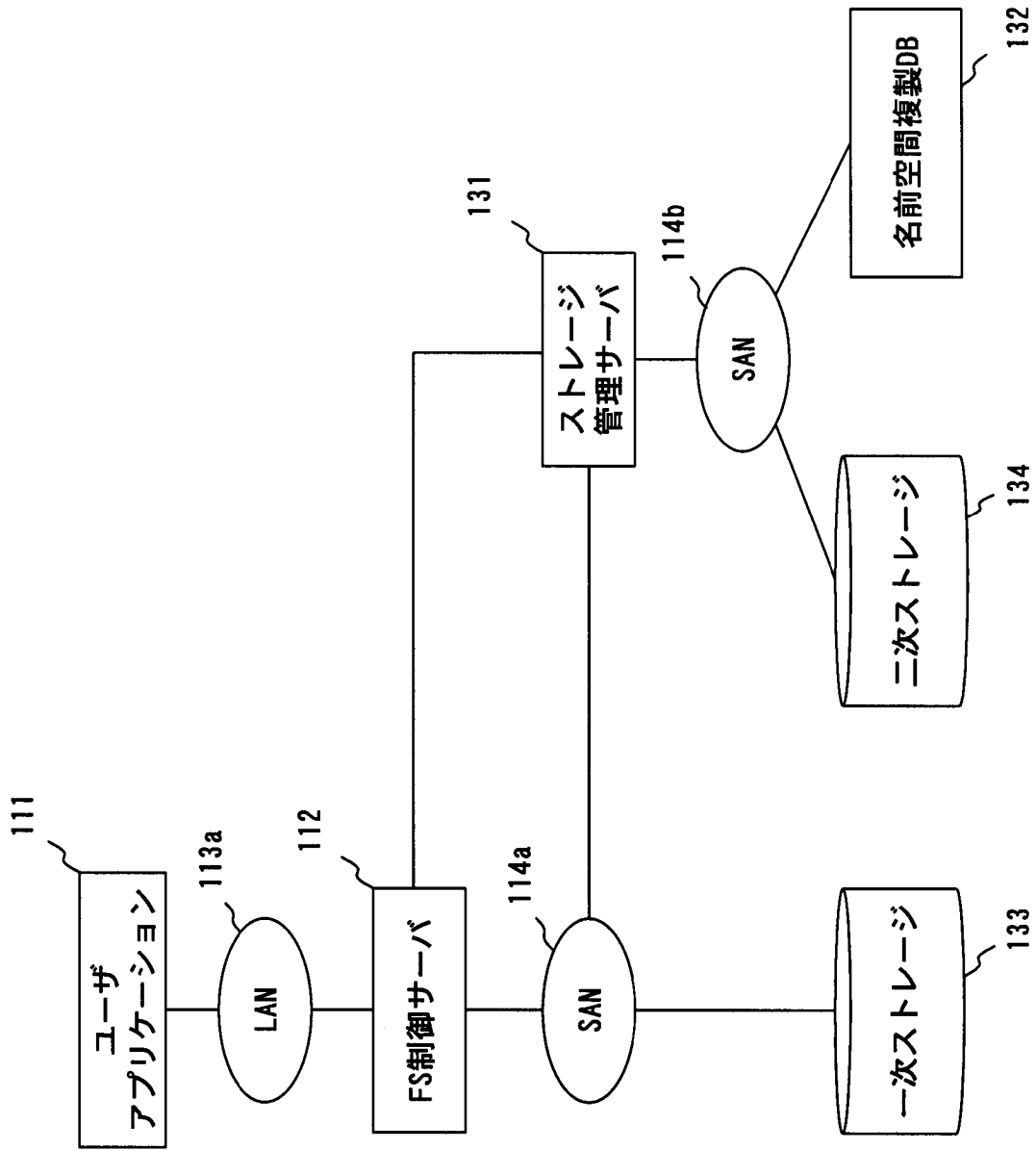
[図5]



[図6]

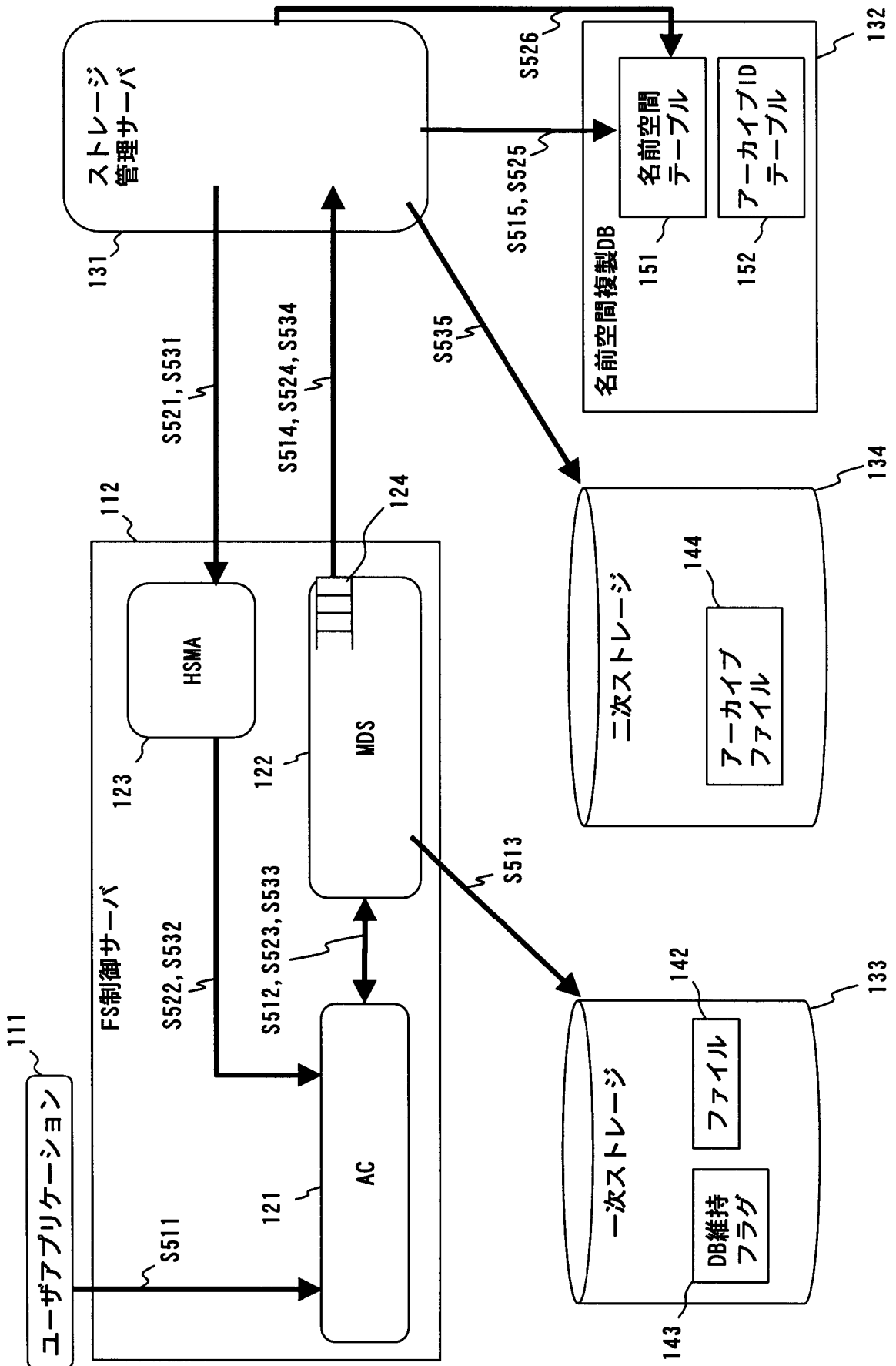


[図7]

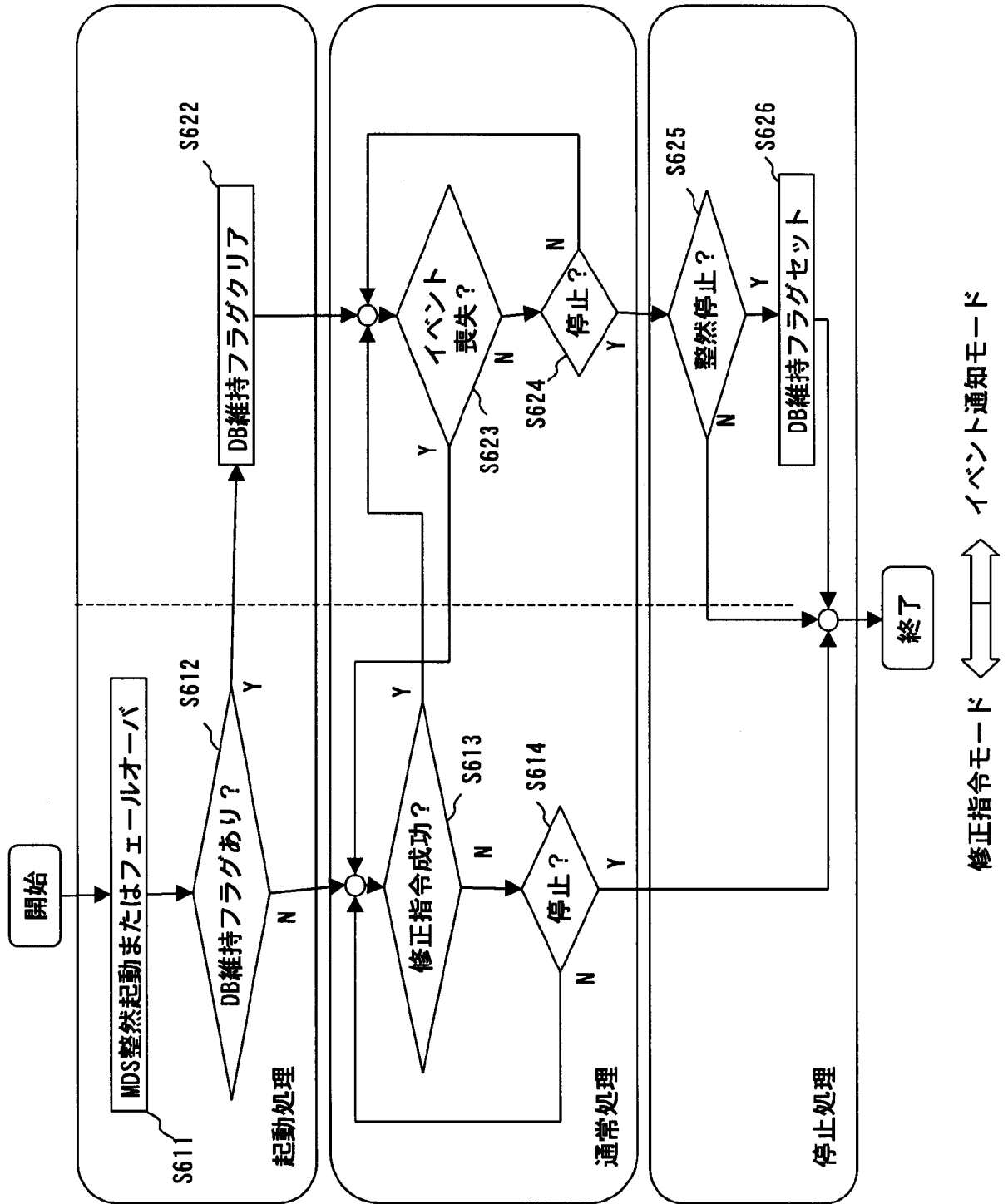




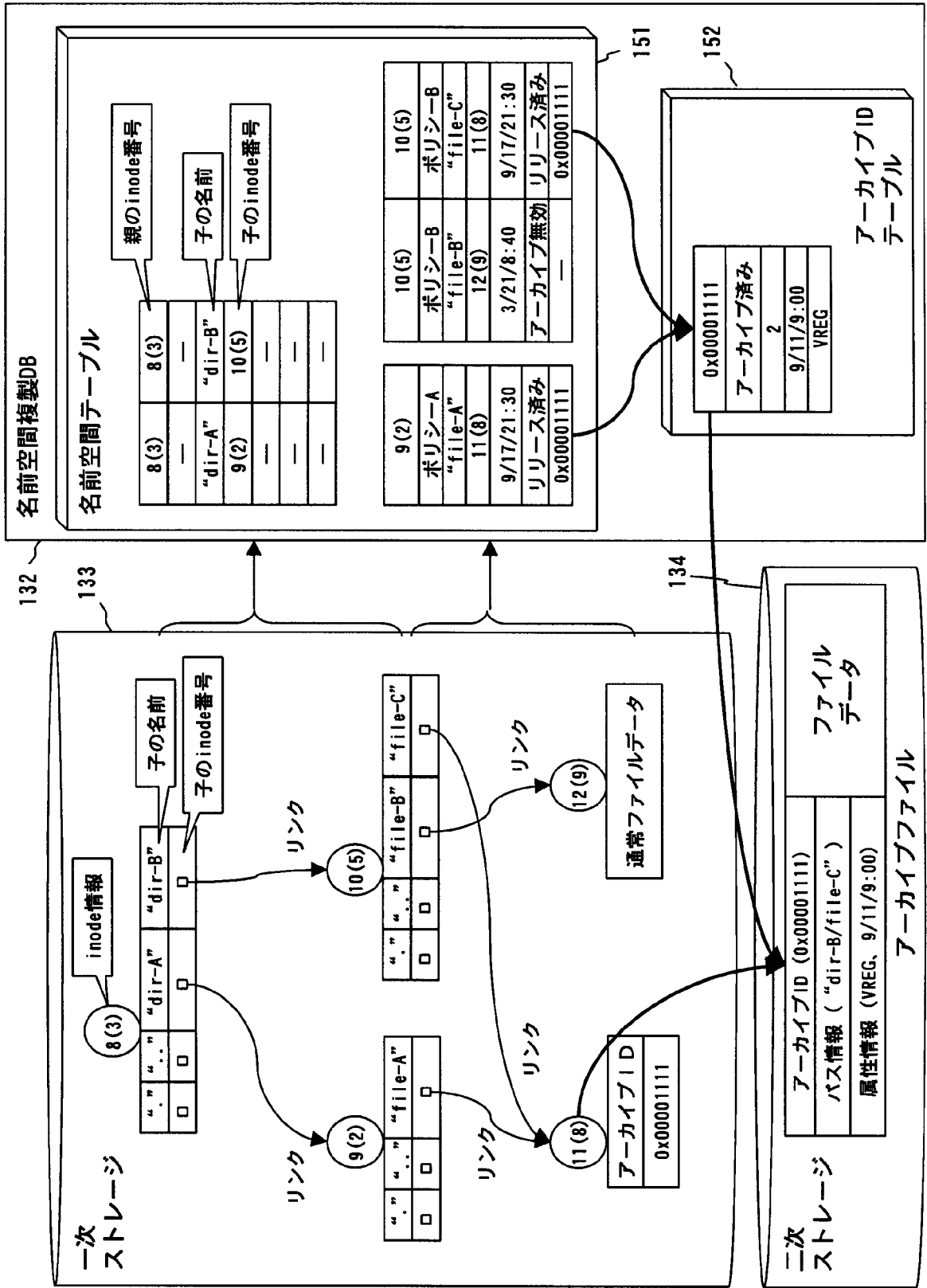
[図8]



[図9]



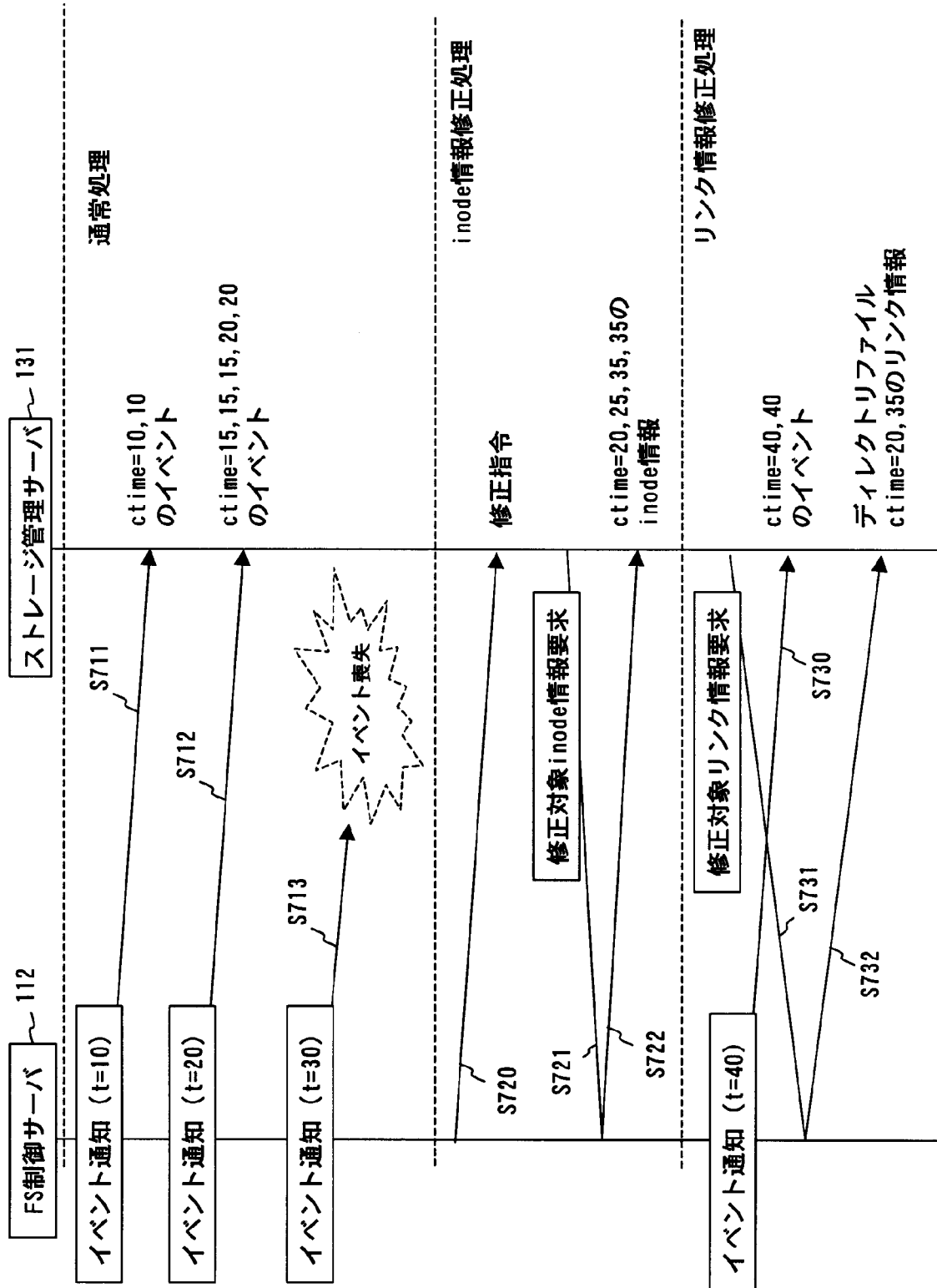
[図10]



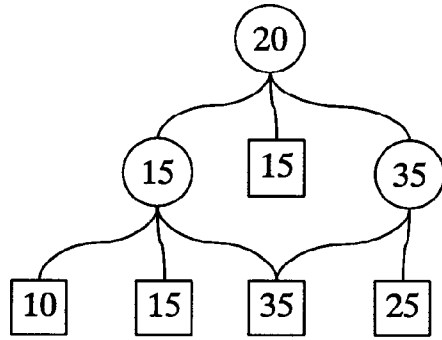
[図11]

	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	○	○	○	○

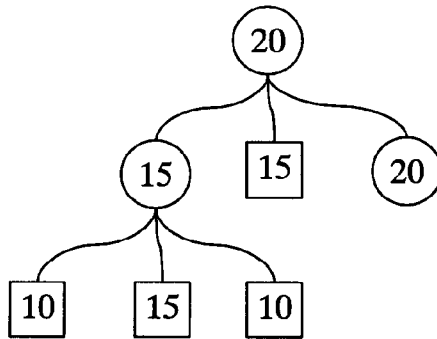
[図12]



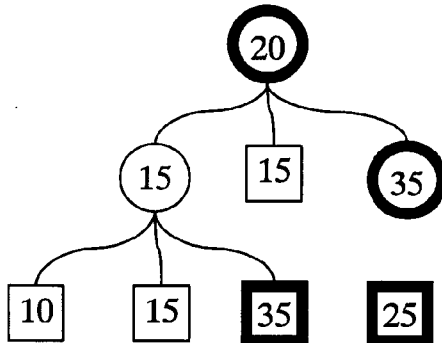
[図13]



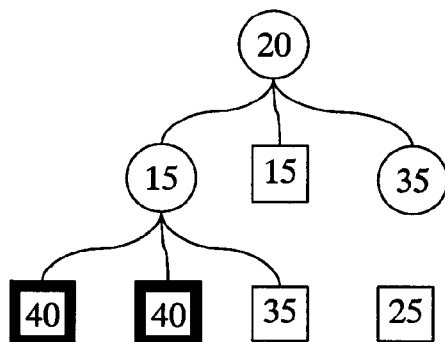
[図14]



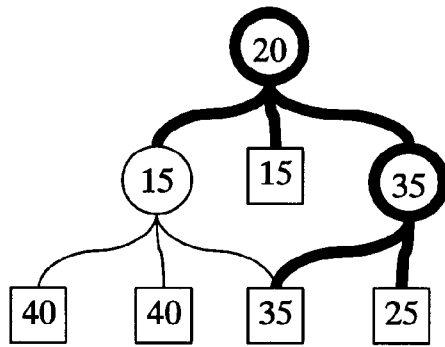
[図15]



[図16]



[図17]



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2006/306284

A. CLASSIFICATION OF SUBJECT MATTER <b>G06F12/00</b> (2006.01)		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) <b>G06F12/00</b> (2006.01)		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Jitsuyo Shinan Koho 1922-1996 Jitsuyo Shinan Toroku Koho 1996-2006 Kokai Jitsuyo Shinan Koho 1971-2006 Toroku Jitsuyo Shinan Koho 1994-2006		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 6-59957 A (Digital Equipment Corp.), 04 March, 1994 (04.03.94), Full text; all drawings & US 5333315 A & GB 2257273 A & DE 4221073 A & FR 2681707 A	1-20
A	JP 2005-196725 A (Hitachi, Ltd.), 21 July, 2005 (21.07.05), Full text; all drawings & US 2005/0125456 A1	1-20
A	JP 2006-39814 A (Hitachi, Ltd.), 09 February, 2006 (09.02.06), Full text; all drawings (Family: none)	1-20
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 26 June, 2006 (26.06.06)		Date of mailing of the international search report 04 July, 2006 (04.07.06)
Name and mailing address of the ISA/ Japanese Patent Office		Authorized officer
Facsimile No.		Telephone No.



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2006/306284

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 1995/029444 A (Sony Corp.), 02 November, 1995 (02.11.95), Full text; all drawings & US 5784646 A & US 6085262 & EP 706129 A1	1-20
A	JP 2003-280950 A (Fujitsu Ltd.), 03 October, 2003 (03.10.03), Full text; all drawings (Family: none)	1-20
A	JP 4-48365 A (Hitachi, Ltd.), 18 February, 1992 (18.02.92), Full text; all drawings (Family: none)	1-20
A	JP 2000-276391 A (Hitachi, Ltd.), 06 October, 2000 (06.10.00), Full text; all drawings (Family: none)	1-20

A. 発明の属する分野の分類 (国際特許分類 (IPC))  
 Int.Cl. G06F12/00(2006.01)

B. 調査を行った分野  
 調査を行った最小限資料 (国際特許分類 (IPC))  
 Int.Cl. G06F12/00(2006.01)

最小限資料以外の資料で調査を行った分野に含まれるもの  
 日本国実用新案公報 1922-1996年  
 日本国公開実用新案公報 1971-2006年  
 日本国実用新案登録公報 1996-2006年  
 日本国登録実用新案公報 1994-2006年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	JP 6-59957 A (デジタル イクイブメント コーポレーション) 1994.03.04, 全文, 全図 & US 5333315 A & GB 2257273 A & DE 4221073 A & FR 2681707 A	1-20
A	JP 2005-196725 A (株式会社日立製作所) 2005.07.21, 全文, 全図 & US 2005/0125456 A1	1-20
A	JP 2006-39814 A (株式会社日立製作所) 2006.02.09, 全文, 全図 (ファミリーなし)	1-20

C欄の続きにも文献が列挙されている。  パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー	の日の後に公表された文献
「A」特に関連のある文献ではなく、一般的技術水準を示すもの	「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの	「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)	「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「O」口頭による開示、使用、展示等に言及する文献	「&」同一パテントファミリー文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願	

国際調査を完了した日 26.06.2006	国際調査報告の発送日 04.07.2006
--------------------------	--------------------------

国際調査機関の名称及びあて先 日本国特許庁 (ISA/J P) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官 (権限のある職員) 相崎 裕恒 電話番号 03-3581-1101 内線 3599	5M	9290
---	--	----	------

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	WO 1995/029444 A (ソニー株式会社) 1995. 11. 02, 全文, 全図 & US 5784646 A & US 6085262 & EP 706129 A1	1-20
A	JP 2003-280950 A (富士通株式会社) 2003. 10. 03, 全文, 全図 (ファミリーなし)	1-20
A	JP 4-48365 A (株式会社日立製作所) 1992. 02. 18, 全文, 全図 (ファミリーなし)	1-20
A	JP 2000-276391 A (株式会社日立製作所) 2000. 10. 06, 全文, 全図 (ファミリーなし)	1-20