

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5018403号
(P5018403)

(45) 発行日 平成24年9月5日(2012.9.5)

(24) 登録日 平成24年6月22日(2012.6.22)

(51) Int.Cl. F 1
G 0 6 F 12/00 (2006.01)
 G 0 6 F 12/00 5 3 1 M
 G 0 6 F 12/00 5 1 0 B

請求項の数 21 (全 18 頁)

(21) 出願番号	特願2007-282582 (P2007-282582)	(73) 特許権者	000004237
(22) 出願日	平成19年10月31日(2007.10.31)		日本電気株式会社
(65) 公開番号	特開2009-110319 (P2009-110319A)		東京都港区芝五丁目7番1号
(43) 公開日	平成21年5月21日(2009.5.21)	(74) 代理人	100088812
審査請求日	平成20年11月12日(2008.11.12)		弁理士 ▲柳▼川 信
		(72) 発明者	大川 人資
			東京都港区芝五丁目7番1号 日本電気株式会社内
		審査官	池田 聡史

最終頁に続く

(54) 【発明の名称】 バックアップシステム、サーバ装置及びそれらに用いるバックアップ方法並びにそのプログラム

(57) 【特許請求の範囲】

【請求項1】

システム上のデータのバックアップを行うバックアップシステムであって、
 前記バックアップ時に同一内容のデータを前記システム上から検索するデータ比較手段と、前記データ比較手段で同一内容のデータが複数存在することが検索された時にそれらのデータがバックアップされているものと見なして当該同一内容のデータをバックアップデータとして扱いかつ当該同一内容のデータのバックアップを抑制するよう制御する制御手段と、前記データ比較手段で検索された前記同一内容のデータ各々の位置情報を記録するカタログ情報記録装置とを有することを特徴とするバックアップシステム。

【請求項2】

前記位置情報は、前記データのファイル名及びパス名を少なくとも含むことを特徴とする請求項1記載のバックアップシステム。

【請求項3】

前記システムが、前記バックアップを制御するサーバ装置と、バックアップ対象の複数のクライアント装置とを含み、

前記データ比較手段で重複が検出されたファイルが予め指定した最低複製数に満たない場合に前記クライアント装置内でバックアップするバックアップ手段を含むことを特徴とする請求項1または請求項2記載のバックアップシステム。

【請求項4】

前記バックアップ手段は、前記重複が検出されたファイルが前記最低複製数より少な

った場合に他のクライアント装置の空き領域に当該ファイルを複製して重複ファイルを作り出すことを特徴とする請求項 3 記載のバックアップシステム。

【請求項 5】

前記カタログ情報記録装置に記録するカタログ情報を生成するカタログ管理手段を含み、

前記カタログ管理手段は、前記ファイル名及び前記パス名とから特定されるファイルについてハッシュ関数で求められたハッシュ値が同じ値となるファイルの組を検出して前記データ比較手段にデータ比較命令を発行することを特徴とする請求項 2 から請求項 4 のいずれか記載のバックアップシステム。

【請求項 6】

前記データ比較手段は、前記サーバ装置と前記複数のクライアント装置とのいずれかに配設したことを特徴とする請求項 3 から請求項 5 のいずれか記載のバックアップシステム。

【請求項 7】

前記サーバ装置は、前記重複ファイルが見つからなかった場合に前記バックアップを行うためのバックアップ記憶装置を含むことを特徴とする請求項 4 から請求項 6 のいずれか記載のバックアップシステム。

【請求項 8】

バックアップ対象の複数のクライアント装置を含むシステム上のデータのバックアップを制御するサーバ装置であって、

前記バックアップ時に同一内容のデータが複数存在することが検索された時にそれらのデータがバックアップされているものと見なして当該同一内容のデータをバックアップデータとして扱いかつ当該同一内容のデータのバックアップを抑止するよう制御する制御手段と、検索された前記同一内容のデータ各々の位置情報を記録するカタログ情報記録装置とを有することを特徴とするサーバ装置。

【請求項 9】

前記位置情報は、前記データのファイル名及びパス名を少なくとも含むことを特徴とする請求項 8 記載のサーバ装置。

【請求項 10】

前記制御手段は、前記同一内容である重複が検出されたファイルが予め指定した最低複製数に満たない場合に前記クライアント装置内でバックアップするバックアップ手段にバックアップ命令を発行することを特徴とする請求項 8 または請求項 9 記載のサーバ装置。

【請求項 11】

前記バックアップ手段は、前記重複が検出されたファイルが前記最低複製数より少なかった場合に他のクライアント装置の空き領域に当該ファイルを複製して重複ファイルを作り出すことを特徴とする請求項 10 記載のサーバ装置。

【請求項 12】

前記カタログ情報記録装置に記録するカタログ情報を生成するカタログ管理手段を含み、

前記カタログ管理手段は、前記ファイル名及び前記パス名とから特定されるファイルについてハッシュ関数で求められたハッシュ値が同じ値となるファイルの組を検出して前記データ比較手段にデータ比較命令を発行することを特徴とする請求項 9 から請求項 11 のいずれか記載のサーバ装置。

【請求項 13】

前記重複ファイルが見つからなかった場合に前記バックアップを行うためのバックアップ記憶装置を含むことを特徴とする請求項 11 または請求項 12 記載のサーバ装置。

【請求項 14】

システム上のデータのバックアップを行うバックアップ方法であって、

前記バックアップ時に同一内容のデータを前記システム上から検索するデータ比較処理と、前記データ比較処理で同一内容のデータが複数存在することが検索された時にそれら

10

20

30

40

50

のデータがバックアップされているものと見なして当該同一内容のデータをバックアップデータとして扱いかつ当該同一内容のデータのバックアップを抑止するよう制御する制御処理と、前記データ比較処理で検索された前記同一内容のデータ各々の位置情報をカタログ情報記録装置に記録する記録処理とを含むことを特徴とするバックアップ方法。

【請求項 15】

前記位置情報が、前記データのファイル名及びパス名を少なくとも含むことを特徴とする請求項 14 記載のバックアップ方法。

【請求項 16】

前記システムが、前記バックアップを制御するサーバ装置と、バックアップ対象の複数のクライアント装置とを含み、

前記クライアント装置が、前記データ比較処理で重複が検出されたファイルが予め指定した最低複製数に満たない場合に前記ファイルのバックアップを行うバックアップ処理を実行することを特徴とする請求項 15 記載のバックアップ方法。

【請求項 17】

前記バックアップ処理において、前記重複が検出されたファイルが前記最低複製数より少なかった場合に他のクライアント装置の空き領域に当該ファイルを複製して重複ファイルを作り出すことを特徴とする請求項 16 記載のバックアップ方法。

【請求項 18】

前記サーバ装置は、前記カタログ情報記録装置に記録するカタログ情報を生成するカタログ管理処理を実行し、

前記カタログ管理処理において、前記ファイル名及び前記パス名とから特定されるファイルについてハッシュ関数で求められたハッシュ値が同じ値となるファイルの組を検出して前記データ比較処理にデータ比較命令を発行することを特徴とする請求項 15 から請求項 17 のいずれか記載のバックアップ方法。

【請求項 19】

前記データ比較処理を、前記サーバ装置と前記複数のクライアント装置とのいずれかが実行することを特徴とする請求項 16 から請求項 18 のいずれか記載のバックアップ方法。

【請求項 20】

前記サーバ装置に、前記重複ファイルが見つからなかった場合に前記バックアップを行うためのバックアップ記憶装置を配設したことを特徴とする請求項 17 から請求項 19 のいずれか記載のバックアップ方法。

【請求項 21】

バックアップ対象の複数のクライアント装置を含むシステム上のデータのバックアップを制御するサーバ装置内において制御手段に実行させるプログラムであって、

前記バックアップ時に同一内容のデータが複数存在することが検索された時にそれらのデータがバックアップされているものと見なして当該同一内容のデータをバックアップデータとして扱いかつ当該同一内容のデータのバックアップを抑止するよう制御する制御処理と、検索された前記同一内容のデータ各々の位置情報をカタログ情報記録装置に記録する記録処理とを含むことを特徴とするプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明はバックアップシステム、サーバ装置及びそれらに用いるバックアップ方法並びにそのプログラムに関し、特にクライアント装置のデータを記録しておくためのバックアップシステムに関する。

【背景技術】

【0002】

本発明に関連するバックアップシステムの一例としては、非特許文献 1, 2 に記載されたシステムがある。このバックアップシステムは、図 12 に示すように、大きくバックア

10

20

30

40

50

ップ装置4とクライアント(#1~#n)5-1~5-nの2つの装置からなる。

【0003】

バックアップ装置4は、スケジュール情報記録装置41と、バックアップジョブ管理部42と、クライアント情報記録装置43と、バックアップカタログ管理部44と、カタログ情報記録装置45と、バックアップ記録装置46とから構成されている。

【0004】

クライアント(#1~#n)5-1~5-nは、データ比較部51-1~51-nと、クライアントバックアップ部52-1~52-nと、バックアップ対象のデータが格納されたユーザデータ領域53-1~53-nとから構成されている。

【0005】

このような構成を有するバックアップシステムは、以下のように動作する。すなわち、バックアップジョブ管理部42は、スケジュール情報記録装置41に格納されたバックアップスケジュールに基づき、クライアント情報記録装置43に記録されたクライアント(#1~#n)5-1~5-nに対してバックアップ命令を発行する。

【0006】

バックアップ命令を受けたクライアント(#1~#n)5-1~5-nは、クライアントバックアップ部52-1~52-nによって、ユーザデータ領域53-1~53-nに含まれるデータをバックアップ記録装置46にバックアップする。この時、データ比較部51-1~51-nによって、既にバックアップ記録装置46に同じデータがバックアップされているか、その有無を判定する。その結果、バックアップされたことのない新しいデータだった場合には、バックアップカタログ管理部44にファイル情報を送信するとともに、バックアップ記録装置46へデータを送信してバックアップする。

【0007】

一方、同じデータが既にバックアップ記憶装置46にバックアップされている場合には、バックアップカタログ管理部44にファイル情報を送信するだけとし、バックアップ記録装置46へのバックアップをスキップする。また、リストアを実行する際には、バックアップ記憶装置46から読み出してリストアする。

【0008】

また、バックアップシステムとしては、下記の特許文献1~5に記載のシステムがある。特許文献1に記載のシステムでは、画像処理装置(クライアント)同士でバックアップを行っており、特許文献2に記載のシステムでは、ネットワークが低負荷時にバックアップを行っており、特許文献3に記載のシステムでは、優先順位に基づいてバックアップを行うことで、不要なバックアップを防止している。

【0009】

特許文献4に記載のシステムでは、バックアップしたファイル(ファイルデータ)と、そのファイルの内容を識別できる属性情報(ファイル情報)とを格納したマスタファイル格納部及び同一性判定部を含んでいる。

【0010】

このシステムでは、まず、ファイルが更新されるか、あるいは新規に作成された場合に、同一性判定部でそれと同じ属性(=内容)のファイルがマスタファイル格納部に格納されているか(バックアップ済みか)どうかを判定する。判定の結果、格納済みであれば処理を終了し、格納されていない場合には、クライアントからデータを取得し、ファイルデータとしてマスタファイル格納部に格納する。その際、ファイルの属性情報をファイル情報として、同様にマスタファイル格納部に格納する。

【0011】

特許文献5に記載のシステムでは、バックアップ時にクライアント上のファイル情報処理部がバックアップ対象ファイルのチェックサム等を含むファイル識別条件を求め、そのファイル識別条件を基に、共有ファイル判別処理部が同一内容のファイルを持つクライアントの数をカウントし、規定数以上であれば共有ファイルとして管理する。規定数以下の場合には、単一ファイルとして管理する。単一ファイルは、そのファイルを持っている各

10

20

30

40

50

クライアント上にあるファイル格納処理部が、そのクライアント上の単一ファイル記憶部にバックアップする。

【0012】

【特許文献1】特開2004-072595号公報

【特許文献2】特開2003-308238号公報

【特許文献3】特開2000-148562号公報

【特許文献4】特開2000-200208号公報

【特許文献5】特開2003-256263号公報

【非特許文献1】EMC Avamar データシート, EMC社 (http://www.emcsoftware.jp/collateral/storage_software/H2568_emc_avamar_ds_ldv.pdf)

10

【非特許文献2】Veritas NetBackup PureDisk™ 6.2 データシート, Symantec社 (http://eval.symantec.com/mktginfo/ja/jp/enterprise/fact_sheets/nbu_pd62_ds_dpro.pdf)

【発明の開示】

【発明が解決しようとする課題】

【0013】

上述した本発明に関連するバックアップ方法では、元のデータが存在する領域とは別の領域に複製を作ることによってデータを保護するため、バックアップするための領域を必ず1世代分用意しなければならないという問題がある。

20

【0014】

また、本発明に関連するバックアップ方法では、重複したファイルが存在する場合でも、それらを検出・記録してリストアに利用する手段を持たない。そのため、本発明に関連するバックアップ方法では、ファイルが破損した際、破損ファイルと同一内容のファイルが別のクライアントにも存在したとしても、それを元に復旧することができないという問題がある。

【0015】

さらに、本発明に関連するバックアップ方法では、バックアップサーバが全てのデータをバックアップ装置に書き出すため、バックアップサーバにトラフィックが集中するという問題がある。特許文献1～3に記載のシステムでも、上記と同様の問題がある。

30

【0016】

特許文献4に記載のシステムでは、バックアップするデータが単一ファイルとなり、各クライアントのバックアップ領域にバックアップされるという問題がある。特許文献5に記載のシステムでは、各クライアントにそれぞれ単一ファイル記憶部を用意しなければならず、そのクライアントの単一ファイルだけがバックアップされるため、各クライアントで単一ファイルの容量をそれぞれ見積る必要がある。

【0017】

そこで、本発明の目的は上記の問題点を解消し、既に複数の重複したファイルが存在するような環境においてバックアップ領域を不要とすることができるバックアップシステム、サーバ装置及びそれらに用いるバックアップ方法並びにそのプログラムを提供することにある。

40

【課題を解決するための手段】

【0018】

本発明によるバックアップシステムは、システム上のデータのバックアップを行うバックアップシステムであって、

前記バックアップ時に同一内容のデータを前記システム上から検索するデータ比較手段と、前記データ比較手段で同一内容のデータが複数存在することが検索された時にそれらのデータがバックアップされているものと見なして当該同一内容のデータをバックアップデータとして扱いかつ当該同一内容のデータのバックアップを抑止するよう制御する制御

50

手段と、前記データ比較手段で検索された前記同一内容のデータ各々の位置情報を記録するカタログ情報記録装置とを備えている。

【0019】

本発明によるサーバ装置は、バックアップ対象の複数のクライアント装置を含むシステム上のデータのバックアップを制御するサーバ装置であって、

前記バックアップ時に同一内容のデータが複数存在することが検索された時にそれらのデータがバックアップされているものと見なして当該同一内容のデータをバックアップデータとして扱いかつ当該同一内容のデータのバックアップを抑止するよう制御する制御手段と、検索された前記同一内容のデータ各々の位置情報を記録するカタログ情報記録装置とを備えている。

10

【0020】

本発明によるバックアップ方法は、システム上のデータのバックアップを行うバックアップ方法であって、

前記バックアップ時に同一内容のデータを前記システム上から検索するデータ比較処理と、前記データ比較処理で同一内容のデータが複数存在することが検索された時にそれらのデータがバックアップされているものと見なして当該同一内容のデータをバックアップデータとして扱いかつ当該同一内容のデータのバックアップを抑止するよう制御する制御処理と、前記データ比較処理で検索された前記同一内容のデータ各々の位置情報をカタログ情報記録装置に記録する記録処理とを含むことを特徴とする。

20

【0021】

本発明によるプログラムは、バックアップ対象の複数のクライアント装置を含むシステム上のデータのバックアップを制御するサーバ装置内において制御手段に実行させるプログラムであって、

前記バックアップ時に同一内容のデータが複数存在することが検索された時にそれらのデータがバックアップされているものと見なして当該同一内容のデータをバックアップデータとして扱いかつ当該同一内容のデータのバックアップを抑止するよう制御する制御処理と、検索された前記同一内容のデータ各々の位置情報をカタログ情報記録装置に記録する記録処理とを含むことを特徴とする。

【発明の効果】

【0022】

本発明は、上記のような構成及び動作とすることで、既に複数の重複したファイルが存在するような環境においてバックアップ領域を不要とすることができるという効果が得られる。

30

【発明を実施するための最良の形態】

【0023】

次に、本発明の実施の形態について図面を参照して説明する。図1は本発明の第1の実施の形態によるバックアップシステムの構成例を示すブロック図である。図1において、本発明の第1の実施の形態によるバックアップシステムは、バックアップを制御する機能を持ったバックアップ管理サーバ1と、バックアップ対象のマシンであるn台のクライアント2-1~2-nとから構成されている。

40

【0024】

バックアップ管理サーバ1は、バックアップされたファイルのカタログ情報を格納するカタログ情報記録装置15と、バックアップカタログ管理部14とを備えている。n台のクライアント2-1~2-nは、バックアップ対象となるユーザデータ25-1~25-nと、データ比較部22-1~22-nと、クライアントバックアップ部23-1~23-nとを備えている。

【0025】

バックアップカタログ管理部14はクライアント2-1~2-nから送信されたファイル名・パス名・ハッシュ値を受信し、カタログ情報記録装置15にカタログ情報を生成する。この時、バックアップカタログ管理部14は、ハッシュ値が同じ値となるファイルの

50

組を検出し、クライアント 2 - 1 ~ 2 - n に対してデータ比較命令を発行する。尚、ハッシュ値の利用は一例であり、他の任意の方法でもよい。

【 0 0 2 6 】

データ比較部 2 2 - 1 ~ 2 2 - n は、バックアップカタログ管理部 1 4 からのデータ比較命令を受信し、指定された 2 つのファイルが同一内容か否かを判定する。同一内容と判定されたファイルは、重複ファイルとしてカタログ情報管理装置 1 5 のカタログ情報に記録される。同一内容でないと判定されたファイルは、単独のファイルとしてカタログ情報に記録される。

【 0 0 2 7 】

クライアントバックアップ部 2 3 - 1 ~ 2 3 - n は、データ比較部 2 2 - 1 ~ 2 2 - n での比較結果で、重複ファイルが最低複製数より少なかったファイルを他のクライアントの空き領域（図示せず）に複製し、重複ファイルを作り出すことでバックアップを実現する。

10

【 0 0 2 8 】

このように、本実施の形態では、データ比較部 2 2 - 1 ~ 2 2 - n で同一内容のデータをシステム上から検索している。この検索において同一内容のものが複数存在した場合、本実施の形態では、それらをバックアップされているものと見なし、それらの位置情報（ファイル名、パス名）をカタログ情報記録装置 1 5 に記録している。また、本実施の形態では、重複しているファイルが予め指定した最低複製数に満たないファイルに対し、クライアントバックアップ部 2 3 - 1 ~ 2 3 - n でバックアップしている。これによって、本実施の形態では、既に複数の重複したファイルが存在するような環境においてバックアップ領域を不要とすることができる。

20

【 0 0 2 9 】

重複排除によってバックアップ領域やデータ領域を削減する方式は多々実施されているが、本実施の形態では、重複を排除するのではなく、バックアップデータとして活用する点が特徴である。既に重複しているファイルは、「既にバックアップされているもの」とみなし、バックアップしない。重複ファイルばかりのデータだった場合には、バックアップに必要な領域が 0 となる。尚、本実施の形態では、データ比較部をクライアント 2 - 1 ~ 2 - n に設けているが、バックアップ管理サーバ 1 に設けることも可能である。

【 0 0 3 0 】

図 2 は本発明の第 2 の実施の形態によるバックアップシステムの構成例を示すブロック図である。図 2 において、本発明の第 2 の実施の形態によるバックアップシステムは、バックアップを制御する機能を持ったバックアップ管理サーバ 1 a と、バックアップ対象のマシンである n 台のクライアント 2 - 1 ~ 2 - n と、外部バックアップ装置 3 とから構成されている。

30

【 0 0 3 1 】

バックアップ管理サーバ 1 a は、スケジュール情報記録装置 1 1 と、クライアント情報記録装置 1 3 と、カタログ情報記録装置 1 5 と、バックアップジョブ管理部 1 2 と、バックアップカタログ管理部 1 4 とを備えている。ここで、スケジュール情報記録装置 1 1 はバックアップ開始時刻を格納し、クライアント情報記録装置 1 3 はバックアップ対象のクライアントの情報を格納し、カタログ情報記録装置 1 5 はバックアップされたファイルのカタログ情報を格納する。また、バックアップジョブ管理部 1 2 及びバックアップカタログ管理部 1 4 は、図示せぬ制御部 [CPU (中央処理装置)] がプログラムを実行することで実現される。

40

【 0 0 3 2 】

n 台のクライアント 2 - 1 ~ 2 - n は、バックアップ対象となるユーザデータ 2 5 - 1 ~ 2 5 - n と、未使用の記録領域である空き領域 2 6 - 1 ~ 2 6 - n と、クライアント情報送信部 2 1 - 1 ~ 2 1 - n と、データ比較部 2 2 - 1 ~ 2 2 - n と、クライアントバックアップ部 2 3 - 1 ~ 2 3 - n と、バックアップデータ維持部 2 4 - 1 ~ 2 4 - n とを備えている。

50

【 0 0 3 3 】

クライアント情報送信部 2 1 - 1 ~ 2 1 - n はユーザデータの情報をバックアップ管理サーバ 1 a に送信し、データ比較部 2 2 - 1 ~ 2 2 - n はファイルの内容を比較する。クライアントバックアップ部 2 3 - 1 ~ 2 3 - n はユーザデータを他のクライアントに複製し、バックアップデータ維持部 2 4 - 1 ~ 2 4 - n はクライアント 2 - 1 ~ 2 - n 上のファイル操作を監視し、バックアップデータを維持する。

【 0 0 3 4 】

上記の各部はそれぞれ次のように動作する。バックアップジョブ管理部 1 2 は、スケジュール情報記録装置 1 1 に格納されたバックアップ時刻に基づき、クライアント情報記録装置 1 3 に記録されたクライアント 2 - 1 ~ 2 - n に対してバックアップ命令を発行する。

10

【 0 0 3 5 】

バックアップカタログ管理部 1 4 はクライアント 2 - 1 ~ 2 - n 上のクライアント情報送信部 2 1 - 1 ~ 2 1 - n によって送信されたファイル名・パス名・ハッシュ値を受信し、カタログ情報記録装置 1 5 にカタログ情報を生成する。この時、バックアップカタログ管理部 1 4 は、ハッシュ値が同じ値となるファイルの組を検出し、クライアント 2 - 1 ~ 2 - n に対してデータ比較命令を発行する。

【 0 0 3 6 】

クライアント情報送信部 2 1 - 1 ~ 2 1 - n は、バックアップジョブ管理部 1 2 からのバックアップ命令を受け、クライアント 2 - 1 ~ 2 - n 内のファイル名、パス名とファイルのハッシュ値を格納したリスト（図 9 参照）を生成し、バックアップカタログ管理部 1 4 に送信する。ハッシュ値が同じファイルが見つかった場合には、後述するデータ比較部 2 2 - 1 ~ 2 2 - n によってファイル内容を比較する。ハッシュ値が同じファイルが無い場合には、単独のファイルとしてカタログ情報に記録する。

20

【 0 0 3 7 】

データ比較部 2 2 - 1 ~ 2 2 - n は、バックアップカタログ管理部 1 4 からのデータ比較命令を受信し、指定された 2 つのファイルが同一内容か否かを判定する。同一内容と判定されたファイルは、重複ファイルとしてカタログ情報管理装置 1 5 のカタログ情報に記録される。同一内容でないと判定されたファイルは、単独のファイルとしてカタログ情報に記録される。

30

【 0 0 3 8 】

クライアントバックアップ部 2 3 - 1 ~ 2 3 - n は、バックアップカタログ管理部 1 4 、及びデータ比較部 2 2 - 1 ~ 2 2 - n の比較結果、重複ファイルが最低複製数より少なかったファイルを他のクライアントの空き領域 2 6 - 1 ~ 2 6 - n に複製し、重複ファイルを作り出すことでバックアップを実現する。

【 0 0 3 9 】

バックアップデータ維持部 2 4 - 1 ~ 2 4 - n は、クライアント 2 - 1 ~ 2 - n 上のファイルアクセスを監視し、重複ファイルとして扱われているファイルの更新や削除を監視し、重複ファイルの数を維持する。クライアント 2 - 1 ~ 2 - n 上でファイルが書込み属性でオープンされると、カタログ情報記録装置 1 5 からそのファイルの重複ファイルの数を取得する。

40

【 0 0 4 0 】

バックアップデータ維持部 2 4 - 1 ~ 2 4 - n は、重複ファイルの数 - 1 が最低複製数より多い場合、カタログ情報を修正して重複ファイルから単独ファイルに変更する。バックアップデータ維持部 2 4 - 1 ~ 2 4 - n は、最低複製数より少なくなる場合、クライアントバックアップ部 2 3 - 1 ~ 2 3 - n によって重複ファイルを新たに作り出し、最低複製数を下回らないようにする。これによって、バックアップデータ維持部 2 4 - 1 ~ 2 4 - n は、重複ファイルの数を維持する。

【 0 0 4 1 】

図 3 及び図 4 は本発明の第 2 の実施の形態によるバックアップシステムの動作を示すシ

50

ーケンスチャートであり、図5は本発明の第2の実施の形態によるバックアップカタログ管理処理を示すフローチャートである。

【0042】

図6は本発明の第2の実施の形態によるデータ比較処理を示すフローチャートであり、図7は本発明の第2の実施の形態によるクライアントバックアップ処理を示すフローチャートであり、図8は本発明の第2の実施の形態によるバックアップデータ維持処理を示すフローチャートである。

【0043】

図9は本発明の第2の実施の形態におけるファイル名とパス名とファイルのハッシュ値とを格納したリストを示す図であり、図10は本発明の第2の実施の形態によるバックアップカタログ管理処理を示す図である。これら図2～図10を参照して本発明の第2の実施の形態によるバックアップシステムの動作について説明する。尚、図3～図8に示す処理のうち、バックアップ管理サーバ1aが実行する処理はバックアップ管理サーバ1a内の制御部[CPU(中央処理装置)]がプログラムを実行することで実現される。

10

【0044】

ここで、図9においては、パス(「クライアント#1-パス#1」、「クライアント#1-パス#2」と、ファイル名(「クライアント#1-ファイル名#1」、「クライアント#1-ファイル名#2」と、ハッシュ値(「ハッシュA」、「ハッシュB」とからなるリストを示している。また、図10においては、ハッシュ値(「ハッシュA」、「ハッシュB」、「ハッシュC」、「ハッシュD」、・・・)で「クライアント名」、「パス」、「ファイル名」を検索する処理を示している。

20

【0045】

バックアップジョブ管理部12にて、最初にスケジュール情報記録装置11に記録されたバックアップ開始時刻を基に、バックアップ開始時刻を決定する(図3のa1)。バックアップジョブ管理部12は、バックアップ開始時刻になると、クライアント情報記録装置13からバックアップ対象クライアントのリストを取得する(図3のa2)。

【0046】

続いて、バックアップジョブ管理部12は、バックアップ対象クライアントのリストに記載されたクライアント2-1～2-nに対し、バックアップ命令を発行する(図3のa3～a5)。バックアップ命令を受けたクライアント2-1～2-nは、クライアント情報送信部21-1～21-nを用いて、重複ファイルの検出を行う(図3のa6, a7)。

30

【0047】

クライアント情報送信部21-1～21-nでは、最初にバックアップ対象であるユーザデータ25-1～25-nに存在するファイル名とパス名との一覧を作成する(図4のb1)。クライアント情報送信部21-1～21-nは、この一覧に挙げられている各ファイルについて、それぞれハッシュ関数でハッシュ値を求める(図4のb2)。

【0048】

さらに、クライアント情報送信部21-1～21-nは、上記のa1の処理で作成したファイル名及びパス名と、上記のa2の処理で作成したハッシュ値とのリスト(図9参照)をバックアップ管理サーバ1aに送信する(図4のb3, a4)。リストを受け取ったバックアップ管理サーバ1aは、バックアップカタログ管理部14によってカタログ情報記録装置15にカタログ情報を作成する(図4のb5)。

40

【0049】

バックアップカタログ管理部14では、まずクライアント情報送信部21-1～21-nから、クライアント2-1～2-n上のファイル名、パス、ハッシュ値のリスト(図9参照)を受信する(図5ステップS1)。バックアップカタログ管理部14は、カタログ情報記録装置15を参照し(図10参照)、クライアント2-1～2-nから得たハッシュ値にエントリが存在するかどうかを確認する(図5ステップS2)。エントリが存在した場合には、同一内容のファイルが存在している可能性を示す。エントリが存在しない場

50

合には、同一内容のファイルがまだ見つからないことを示す。

【 0 0 5 0 】

エントリが存在した場合、バックアップカタログ管理部 1 4 は、カタログ情報から同一のハッシュ値を持つエントリを 1 つピックアップする (図 5 ステップ S 6)。この時、バックアップカタログ管理部 1 4 は、ネットワーク上の距離が近いものを優先してピックアップする。

【 0 0 5 1 】

さらに、バックアップカタログ管理部 1 4 は、ステップ S 1 の処理で得た情報が指すファイルと、ステップ S 6 の処理でピックアップしたファイルとの内容をデータ比較部 2 2 - 1 ~ 2 2 - n で比較する (図 5 ステップ S 7)。

10

【 0 0 5 2 】

データ比較部 2 2 - 1 ~ 2 2 - n では、バックアップカタログ管理部 1 4 から、自身が持つ比較対象のファイルのパス名、ファイル名、比較相手のクライアント名、そのクライアント上でのパス名、ファイル名を受信する (図 6 ステップ S 1 1)。データ比較部 2 2 - 1 ~ 2 2 - n は、比較相手のクライアントのデータ比較部から相手側のファイル内容を取得し (図 6 ステップ S 1 2)、自身が持つ比較対象のファイルの内容と比較する (図 6 ステップ S 1 3)。

【 0 0 5 3 】

データ比較部 2 2 - 1 ~ 2 2 - n は、その比較結果を確認し (図 6 ステップ S 1 4)、ファイル内容が同一だった場合、重複ファイルとして、バックアップカタログ管理部 1 4 に回答する (図 6 ステップ S 1 5)。データ比較部 2 2 - 1 ~ 2 2 - n は、内容が一致しない場合、別ファイルとしてバックアップカタログ管理部 1 4 に回答する (図 6 ステップ S 1 6)。

20

【 0 0 5 4 】

バックアップカタログ管理部 1 4 は、データ比較部 2 2 - 1 ~ 2 2 - n からの回答を受け、内容が同一のファイルだった場合 (図 5 ステップ S 8)、重複ファイルとしてカタログ情報に追加する (図 5 ステップ S 9)。バックアップカタログ管理部 1 4 は、内容が同一ではなかった場合 (図 5 ステップ S 8)、単独の新しいファイルとして、カタログ情報に追加する (図 5 ステップ S 1 0)。

【 0 0 5 5 】

バックアップカタログ管理部 1 4 は、ステップ S 2 の処理にて、エントリが見つからなかった場合、ステップ S 1 の処理で得たハッシュ値、ファイル名、パス名をカタログ情報記録装置 1 5 上のカタログ情報に登録する (図 5 ステップ S 3)。

30

【 0 0 5 6 】

バックアップカタログ管理部 1 4 は、クライアント 2 - 1 ~ 2 - n 上の全ファイルの情報を処理するまで上記の処理を繰り返し (図 5 ステップ S 4)、最後に、重複ファイルの数が予め設定された最低複製数以下になったエントリをカタログ情報からピックアップし、クライアントバックアップ部 2 3 - 1 ~ 2 3 - n でバックアップする (図 5 ステップ S 5)。この場合、バックアップ先としては、クライアント情報記録装置 1 3 から、空き領域 2 6 - 1 ~ 2 6 - n のサイズや、ネットワーク上での距離を考慮して決定する。

40

【 0 0 5 7 】

バックアップカタログ管理部 1 4 からバックアップ命令を受けた送信側のクライアントバックアップ部 2 3 - 1 ~ 2 3 - n は、バックアップ命令に続いてバックアップ対象のファイルのパス名、ファイル名、バックアップ先のクライアント名を受信する (図 7 ステップ S 2 1)。送信側のクライアントバックアップ部 2 3 - 1 ~ 2 3 - n は、バックアップ先のクライアントバックアップ部に接続し、バックアップ対象であるファイルの内容を送信する (図 7 ステップ S 2 2)。

【 0 0 5 8 】

受信側のクライアントバックアップ部 2 3 - 1 ~ 2 3 - n では、受け取ったバックアップデータを自身の空き領域 2 6 - 1 ~ 2 6 - n に保存する (図 7 ステップ S 3 1)。この

50

場合、バックアップ先としては、外部バックアップ装置 3 を利用することも可能である。

【 0 0 5 9 】

受信側のクライアントバックアップ部 2 3 - 1 ~ 2 3 - n は、空き領域 2 6 - 1 ~ 2 6 - n に保存したバックアップデータのパス名、ファイル名、ハッシュ値をバックアップ管理サーバ 1 に送信する (図 7 ステップ S 3 2)。バックアップ管理サーバ 1 は、バックアップカタログ管理部 1 4 によって、カタログ情報記録装置 1 5 にカタログ情報を作成する。

【 0 0 6 0 】

重複ファイルとしてカタログ情報に登録され、バックアップデータとして扱われているファイルは、ユーザのファイル操作によって変更、削除される可能性がある。このため、バックアップデータ維持部 2 4 - 1 ~ 2 4 - n では、最低複製数を維持する。バックアップデータ維持部 2 4 - 1 ~ 2 4 - n は、クライアント 2 - 1 ~ 2 - n 上のファイルが、読み出し以外の操作 (更新、削除等) を行うためにオープンされた場合に実行する。

【 0 0 6 1 】

まず、バックアップデータ維持部 2 4 - 1 ~ 2 4 - n は、バックアップ管理サーバ 1 上のカタログ情報記録装置 1 5 から、オープンされたファイルが何個の重複ファイルを持つか、問い合わせる (図 8 ステップ S 4 1)。バックアップデータ維持部 2 4 - 1 ~ 2 4 - n は、問い合わせの結果で得た重複ファイル数と最低複製数とを比較する (図 8 ステップ S 4 2)。

【 0 0 6 2 】

バックアップデータ維持部 2 4 - 1 ~ 2 4 - n は、重複ファイル数が最低複製数以上だった場合、バックアップ管理サーバ 1 上のカタログ情報記録装置 1 5 から、オープンされたファイルのエントリを削除する (図 8 ステップ S 4 3)。バックアップデータ維持部 2 4 - 1 ~ 2 4 - n は、ファイルへの操作が変更・移動だった場合、新しいパス名とファイル名、及びハッシュ値をバックアップ管理サーバ 1 に送信する (図 8 ステップ S 4 4)。

【 0 0 6 3 】

バックアップ管理サーバ 1 は、バックアップカタログ管理部 1 4 によって、カタログ情報記録装置 1 5 にカタログ情報を作成する。重複ファイル数が最低複製数未満だった場合には、クライアントバックアップ部 2 3 - 1 ~ 2 3 - n でバックアップする (図 8 ステップ S 4 5)。バックアップ先はクライアント情報記録装置 1 3 から、空き領域 2 6 - 1 ~ 2 6 - n のサイズや、ネットワーク上での距離を考慮して決定する。

【 0 0 6 4 】

このように、本実施の形態では、重複したファイルを検出し、それらを既に冗長化された状態にあるものとみなしてバックアップしないため、既に複数の重複したファイルが存在するような環境において、バックアップ領域を不要とすることができる。

【 0 0 6 5 】

また、本実施の形態では、重複ファイルが見つからなかった場合のバックアップを、クライアント 2 - 1 ~ 2 - n 上の空き領域 2 6 - 1 ~ 2 6 - n に取るように構成しているため、バックアップ管理サーバ 1 にバックアップ記憶装置が不要となる。

【 0 0 6 6 】

さらに、本実施の形態では、バックアップ時に重複ファイルを検出するとともに、そのファイルの更新・削除を監視し、記録・維持するため、ファイルが破損して復旧が必要になった場合、他の重複したファイルから復旧することができる。

【 0 0 6 7 】

さらに、本実施の形態では、バックアップ時に重複ファイルを検出するとともに、そのファイルの更新・削除を監視し、記録・維持するため、重複したファイルが現時点で全システム上に何個存在するのか、同名のファイルに幾つのバージョンが存在するのかを確認することができる。

【 0 0 6 8 】

図 1 1 は本発明の第 3 の実施の形態によるバックアップシステムの構成例を示すブロッ

10

20

30

40

50

ク図である。図 1 1 において、本発明の第 3 の実施の形態によるバックアップシステムは、バックアップを制御する機能を持ったバックアップ管理サーバ 1 b と、バックアップ対象のマシンである n 台のクライアント 2 a - 1 ~ 2 a - n とから構成されている。

【 0 0 6 9 】

バックアップ管理サーバ 1 b は、バックアップ開始時刻を格納するスケジュール情報記録装置 1 1 と、クライアント情報記録装置 1 3 と、カタログ情報記録装置 1 5 と、バックアップジョブ管理部 1 2 と、バックアップカタログ管理部 1 4 と、バックアップ記憶装置 1 6 と、データ比較部 1 7 とを備えている。ここで、クライアント情報記録装置 1 3 はバックアップ対象のクライアントの情報を格納し、カタログ情報記録装置 1 5 はバックアップされたファイルのカタログ情報を格納する。

10

【 0 0 7 0 】

n 台のクライアント 2 a - 1 ~ 2 a - n は、バックアップ対象となるユーザデータ 2 5 - 1 ~ 2 5 - n と、クライアント情報送信部 2 1 - 1 ~ 2 1 - n と、クライアントバックアップ部 2 3 - 1 ~ 2 3 - n と、バックアップデータ維持部 2 4 - 1 ~ 2 4 - n とを備えている。

【 0 0 7 1 】

クライアント情報送信部 2 1 - 1 ~ 2 1 - n は、ユーザデータの情報をバックアップ管理サーバ 1 b に送信し、クライアントバックアップ部 2 3 - 1 ~ 2 3 - n はユーザデータを他のクライアントに複製し、バックアップデータ維持部 2 4 - 1 ~ 2 4 - n はクライアント 2 a - 1 ~ 2 a - n 上のファイル操作を監視し、バックアップデータを維持する。図 1 1 においては、ファイル A , C , D が重複ファイルであり、ファイル B が単一ファイルである。

20

【 0 0 7 2 】

これらの各部はそれぞれ次のように動作する。バックアップジョブ管理部 1 2 は、スケジュール情報記録装置 1 1 に格納されたバックアップ時刻に基づき、クライアント情報記録装置 1 3 に記録されたクライアントに対してバックアップ命令を発行する。

【 0 0 7 3 】

クライアント情報送信部 2 1 - 1 ~ 2 1 - n は、バックアップジョブ管理部 1 2 からのバックアップ命令を受け、クライアント 2 a - 1 ~ 2 a - n 内のファイル名、パス名、ファイルのハッシュ値を格納したリスト（図 9 参照）を生成し、バックアップ管理部 1 4 に送信する。同時に、クライアントバックアップ部 2 3 - 1 ~ 2 3 - n は、ユーザデータ 2 5 - 1 ~ 2 5 - n をバックアップ管理サーバ 1 b 上のバックアップ記憶装置 1 6 に送信する。

30

【 0 0 7 4 】

バックアップカタログ管理部 1 4 は、クライアント 2 a - 1 ~ 2 a - n 上のクライアント情報送信部 2 1 - 1 ~ 2 1 - n によって送信されたファイル名・パス名・ハッシュ値を受信し、カタログ情報記録装置 1 5 にカタログ情報を生成する。この時、バックアップカタログ管理部 1 4 は、ハッシュ値が同じ値となるファイルの組を検出し、データ比較部 1 7 で双方のデータ比較を実行する。

【 0 0 7 5 】

同一内容と判定されたファイルは、重複ファイルとしてカタログ情報管理装置 1 5 のカタログ情報に記録し、バックアップ記憶装置 1 6 からバックアップしたデータを削除する。

40

【 0 0 7 6 】

バックアップデータ維持部 2 4 - 1 ~ 2 4 - n は、クライアント 2 a - 1 ~ 2 a - n 上のファイルアクセスを監視し、重複ファイルとして扱われているファイルの更新や削除を監視し、重複ファイルの数を維持する。バックアップデータ維持部 2 4 - 1 ~ 2 4 - n は、クライアント 2 a - 1 ~ 2 a - n 上でファイルが書き込み属性でオープンされると、カタログ情報記録装置 1 5 からそのファイルの重複ファイルの数を取得する。

【 0 0 7 7 】

50

バックアップデータ維持部 24 - 1 ~ 24 - n は、重複ファイルの数 - 1 が最低複製数より多い場合、カタログ情報を修正して重複ファイルから単独ファイルに変更する。バックアップデータ維持部 24 - 1 ~ 24 - n は、最低複製数より少なくなる場合、クライアントバックアップ部 23 - 1 ~ 23 - n によって重複ファイルを新たに作り出し、最低複製数を下回らないようにする。これによって、バックアップデータ維持部 24 - 1 ~ 24 - n は、重複ファイルの数を維持する。

【0078】

このように、本実施の形態では、ファイルのデータ比較をクライアント 2a - 1 ~ 2a - n 上ではなく、バックアップ管理サーバ 1b 上で行うように構成することによって、上記の本発明の第 2 の実施の形態と比較し、クライアント 2a - 1 ~ 2a - n 上の負荷を削減することができる。

10

【0079】

また、本実施の形態では、さらに重複ファイルが見つからなかった場合のバックアップを、クライアント 2a - 1 ~ 2a - n 上の空き領域ではなく、バックアップ管理サーバ 1b 上のバックアップ記憶装置 16 にバックアップするように構成しているため、将来、バックアップ量が増加した場合でも、バックアップ管理サーバ 1b のバックアップ記憶装置 16 の増強だけで対処することができる。

【産業上の利用可能性】

【0080】

本発明は、オフィスや事業場内等、多数の PC (Personal Computer) が存在する環境やシンクライアント環境でのバックアップといった用途に適用することができる。

20

【図面の簡単な説明】

【0081】

【図 1】本発明の第 1 の実施の形態によるバックアップシステムの構成例を示すブロック図である。

【図 2】本発明の第 2 の実施の形態によるバックアップシステムの構成例を示すブロック図である。

【図 3】本発明の第 2 の実施の形態によるバックアップシステムの動作を示すシーケンスチャートである。

30

【図 4】本発明の第 2 の実施の形態によるバックアップシステムの動作を示すシーケンスチャートである。

【図 5】本発明の第 2 の実施の形態によるバックアップカタログ管理処理を示すフローチャートである。

【図 6】本発明の第 2 の実施の形態によるデータ比較処理を示すフローチャートである。

【図 7】本発明の第 2 の実施の形態によるクライアントバックアップ処理を示すフローチャートである。

【図 8】本発明の第 2 の実施の形態によるバックアップデータ維持処理を示すフローチャートである。

【図 9】本発明の第 2 の実施の形態におけるファイル名とパス名とファイルのハッシュ値とを格納したリストを示す図である。

40

【図 10】本発明の第 2 の実施の形態によるバックアップカタログ管理処理を示す図である。

【図 11】本発明の第 3 の実施の形態によるバックアップシステムの構成例を示すブロック図である。

【図 12】本発明に関連するバックアップシステムの構成例を示すブロック図である。

【符号の説明】

【0082】

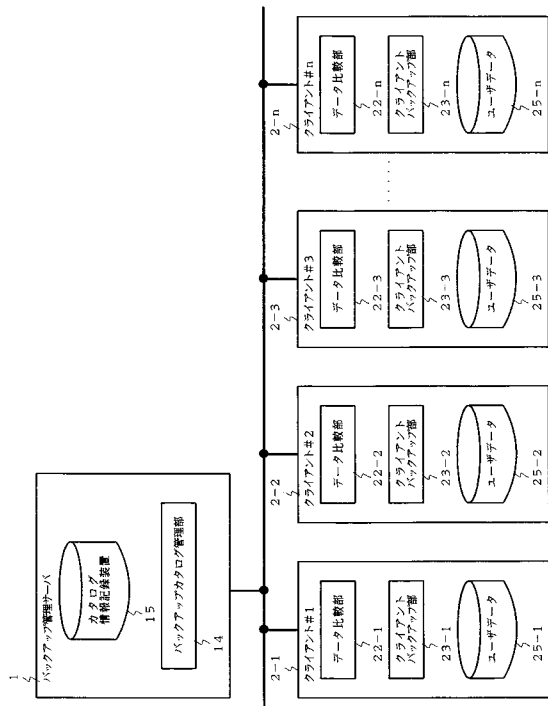
1, 1a, 1b バックアップ管理サーバ
2 - 1 ~ 2 - n,

50

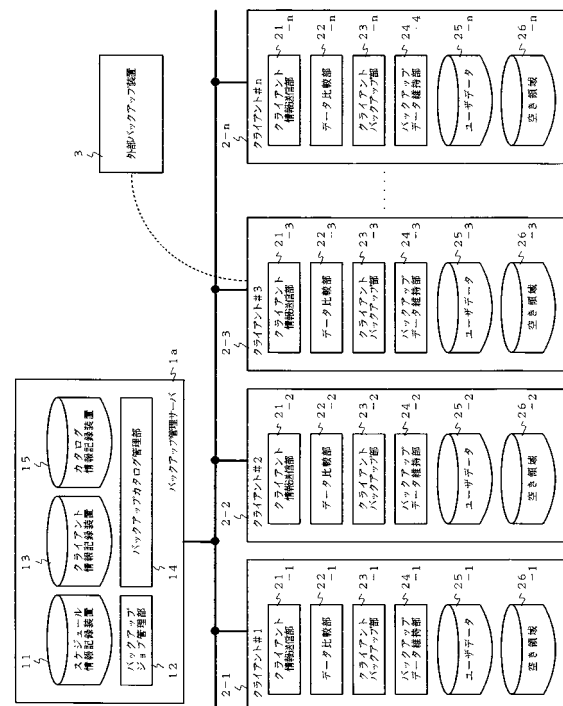
- 2 a - 1 ~ 2 a - n クライアント
- 3 外部バックアップ装置
- 1 1 スケジュール情報記録装置
- 1 2 バックアップジョブ管理部
- 1 3 クライアント情報記録装置
- 1 4 バックアップカタログ管理部
- 1 5 カタログ情報記録装置
- 1 6 バックアップ記憶装置

- 1 7 ,
- 2 2 - 1 ~ 2 2 - n データ比較部
- 2 1 - 1 ~ 2 1 - n クライアント情報送信部
- 2 3 - 1 ~ 2 3 - n クライアントバックアップ部
- 2 4 - 1 ~ 2 4 - n バックアップデータ維持部
- 2 5 - 1 ~ 2 5 - n ユーザデータ
- 2 6 - 1 ~ 2 6 - n 空き領域

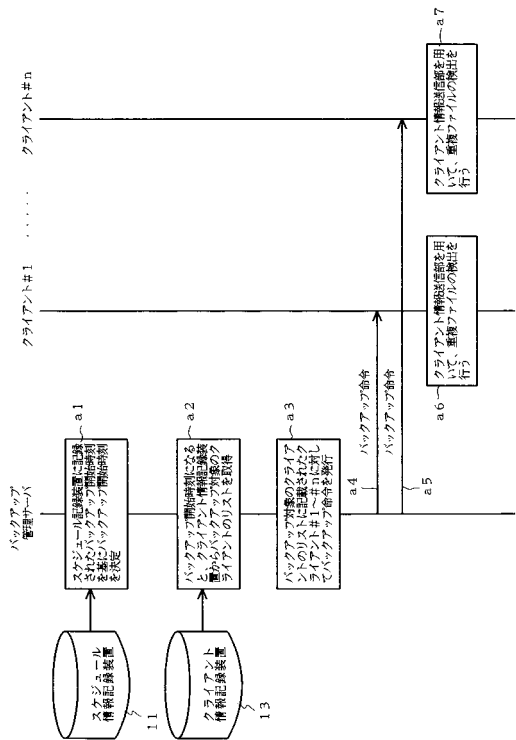
【 図 1 】



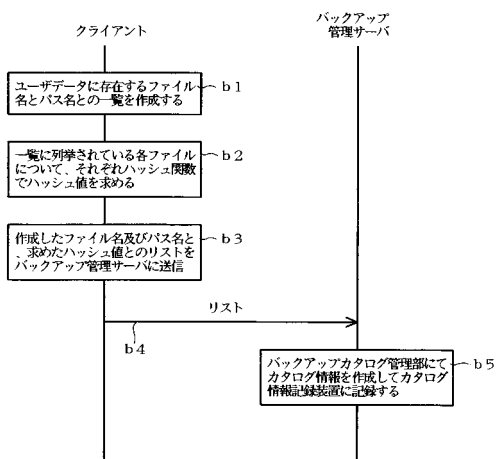
【 図 2 】



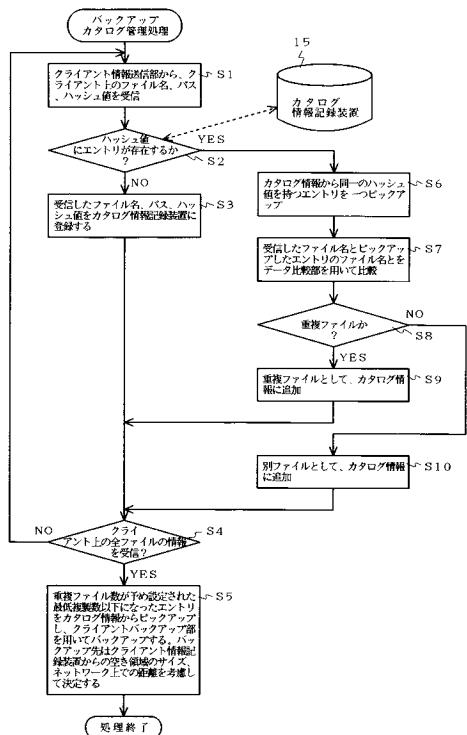
【図3】



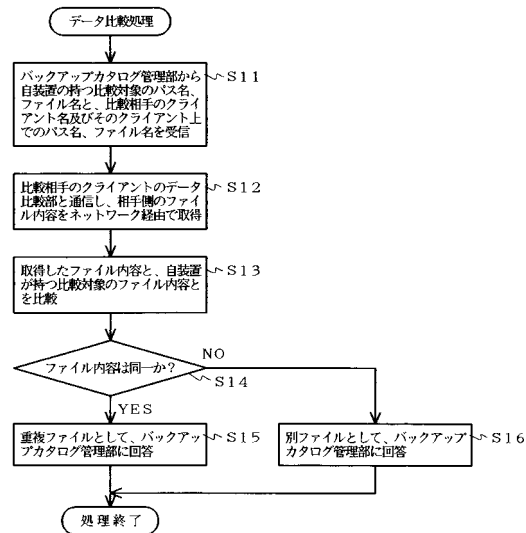
【図4】



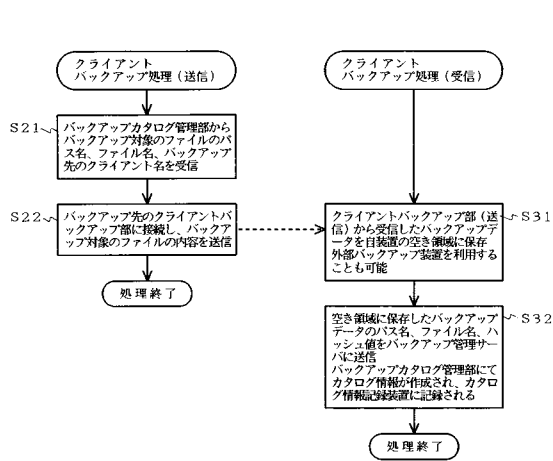
【図5】



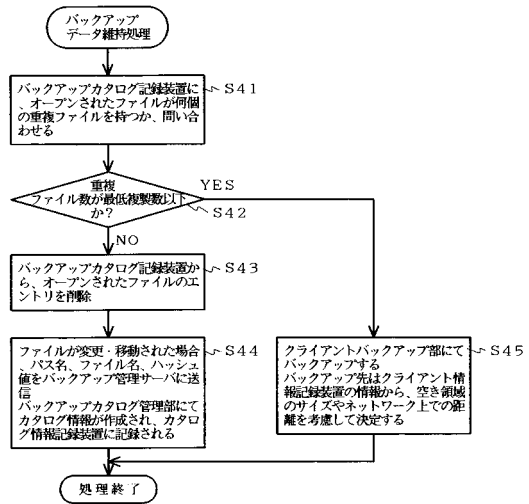
【図6】



【図7】



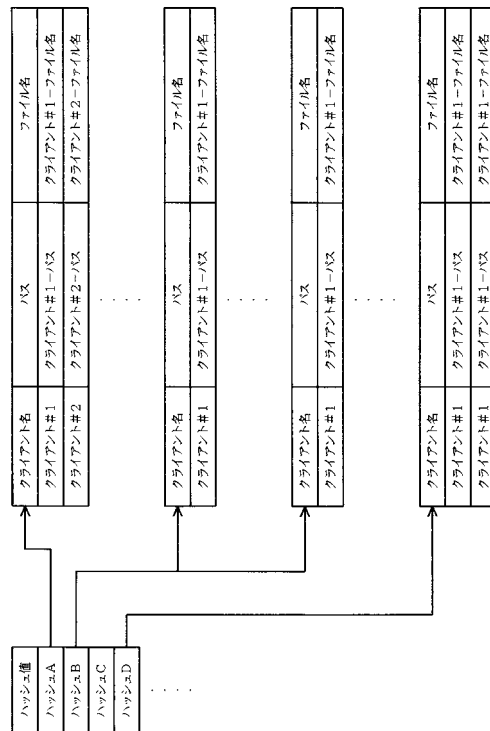
【図8】



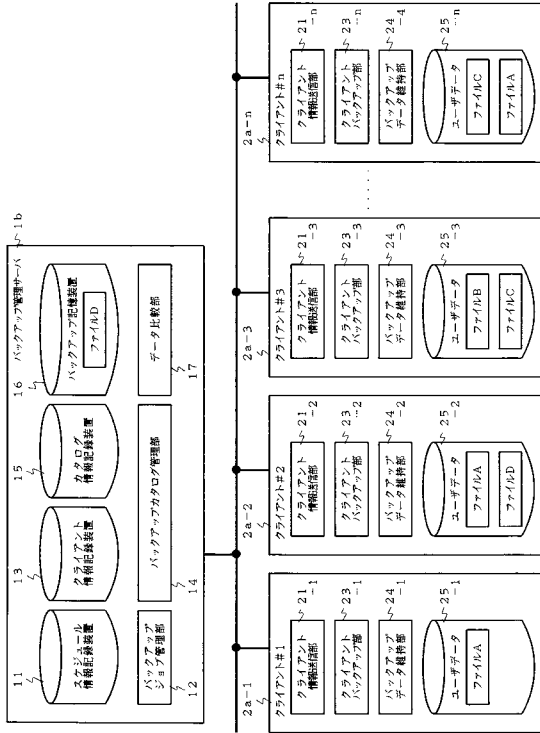
【図9】

パス	ファイル名	ハッシュ値
クライアント#1-パス#1	クライアント#1-ファイル名#1	ハッシュA
クライアント#1-パス#2	クライアント#1-ファイル名#2	ハッシュB

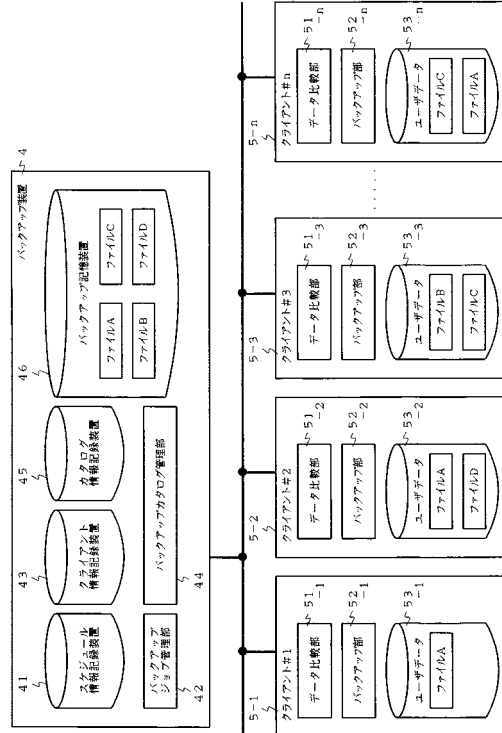
【図10】



【図 1 1】



【図 1 2】



フロントページの続き

- (56)参考文献 特開2000-293420(JP,A)
特開2006-350470(JP,A)
特開2002-236605(JP,A)
特開2002-099454(JP,A)
特開2002-278818(JP,A)
特開2004-302573(JP,A)
特開平10-124364(JP,A)
特開平6-332776(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F 12/00