(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5452765号
(P5452765)

(45)発行日　平成26年3月26日(2014.3.26)　(24)登録日　平成26年1月10日(2014.1.10)

(51)Int.Cl.　　F I
G06F 12/00 (2006.01)　　G06F 12/00 531R
G06F 3/06 (2006.01)　　G06F 3/06 306H
G06F 3/06 306K
G06F 3/06 304E
G06F 3/06 301X

請求項の数 15　(全 53 頁)

(21)出願番号　特願2013-501477 (P2013-501477)
(86)(22)出願日　平成22年12月14日 (2010.12.14)
(65)公表番号　特表2013-532314 (P2013-532314A)
(43)公表日　平成25年8月15日 (2013.8.15)
(86)国際出願番号　PCT/JP2010/007254
(87)国際公開番号　WO2012/081050
(87)国際公開日　平成24年6月21日 (2012.6.21)
審査請求日　平成25年1月11日 (2013.1.11)

(73)特許権者　000005108
株式会社日立製作所
東京都千代田区丸の内一丁目６番６号
(74)代理人　110000176
一色国際特許業務法人
(72)発明者　雜賀　信之
神奈川県小田原市中里３２２番２号　株式会社日立製作所ＲＡＩＤシステム事業部内

審査官　池田　聡史

最終頁に続く

(54)【発明の名称】情報処理システムにおける障害復旧方法、及び情報処理システム

(57)【特許請求の範囲】
【請求項１】
第１ファイルシステムを有し、データＩ／Ｏ要求を受け付ける第１サーバ装置と、
第２ファイルシステムを有し、前記第１サーバ装置に通信可能に接続された第２サーバ装置と、を備え、
前記第１サーバ装置は、前記データＩ／Ｏ要求の対象であるファイルのデータを第１ストレージ装置に記憶し、
前記第２サーバ装置は、前記データＩ／Ｏ要求の対象であるファイルのデータを第２ストレージ装置に記憶し、
前記第１サーバ装置が、前記第１ストレージ装置に記憶しているファイルのデータを前記第２サーバ装置に送信し、
前記第２サーバ装置が、前記第１サーバ装置から送られてくる前記データを前記第２ストレージ装置に記憶している情報処理システムにおける障害の復旧方法であって、
前記第２サーバ装置は、障害からの復旧に際して前記第１サーバ装置が前記データＩ／Ｏ要求の受け付けを開始するのに先立ち、前記第２ストレージ装置に記憶しているディレクトリイメージのうち最上位のディレクトリから所定の下位階層までのディレクトリイメージを、前記第１サーバ装置に送信し、
前記第１サーバ装置は、前記第２サーバ装置から送られてくる前記ディレクトリイメージを前記第１ストレージ装置に復元した後、前記データＩ／Ｏ要求の受け付けを再開し、
前記第１サーバ装置は、前記データＩ／Ｏ要求の受け付け再開後において、受け付けた

前記データＩ／Ｏ要求を処理するために必要となるディレクトリイメージが前記第１ストレージ装置に復元されていない場合には、前記ディレクトリイメージを前記第２サーバ装置に要求し、
前記第２サーバ装置は、前記第１サーバ装置から送られてくる前記要求に応じて前記ディレクトリイメージを前記第２ストレージ装置から読み出して前記第１サーバ装置に送信し、
前記第１サーバ装置は、前記第２ストレージ装置から送られてくる前記ディレクトリイメージに基づき前記データＩ／Ｏ要求を処理するとともに、前記ディレクトリイメージを前記第１ストレージ装置に復元する
情報処理システムにおける障害復旧方法。
【請求項２】
請求項１に記載の情報処理システムにおける障害復旧方法であって、
前記第２サーバ装置は、前記第１サーバ装置から前記要求が送られてきた場合、前記要求の対象である前記ディレクトリイメージを前記第２ストレージ装置から読み出して前記第１サーバ装置に送信するとともに、所定の選出方法に従って選出した、前記ディレクトリイメージとは異なる追加のディレクトリイメージを、前記第２ストレージ装置から読み出して前記第１サーバ装置に送信し、
前記第１サーバ装置は、前記第２サーバ装置から送られてくる前記ディレクトリイメージに基づき前記データＩ／Ｏ要求を処理するとともに、前記ディレクトリイメージ及び前記第２サーバ装置から送られてくる前記追加のディレクトリイメージを、前記第１ストレージ装置に復元する
情報処理システムにおける障害復旧方法。
【請求項３】
請求項２に記載の情報処理システムにおける障害復旧方法であって、
前記第２サーバ装置は、前記第１サーバ装置が受け付けた前記データＩ／Ｏ要求が所定の条件を満たす場合に、所定の選出方法に従って前記追加のディレクトリイメージを選出する
情報処理システムにおける障害復旧方法。
【請求項４】
請求項３に記載の情報処理システムにおける障害復旧方法であって、
前記所定の条件は、
前記データＩ／Ｏ要求の対象になっているファイルのデータサイズが、現在から遡って所定時間内に発生したデータＩ／Ｏ要求が対象としていたファイルのデータサイズの平均値よりも小さいこと、
前記データＩ／Ｏ要求の対象になっているファイルのデータサイズが予め設定された閾値よりも小さいこと
のうちの少なくともいずれかである
情報処理システムにおける障害復旧方法。
【請求項５】
請求項３に記載の情報処理システムにおける障害復旧方法であって、
前記所定の選出方法は、
前記第１ストレージ装置に既に復元されているディレクトリの配下に存在するファイルのメタデータ及び／又は実体を前記追加のディレクトリイメージとする方法、
前記第１ストレージ装置に既に復元されているディレクトリの配下に存在するディレクトリのメタデータを前記追加のディレクトリイメージとする方法、
前記障害が発生する前に前記第１ストレージ装置に実体が格納されていたファイルの実体を前記追加のディレクトリイメージとする方法、
重要度が高く設定されているファイルのメタデータ及び／又は実体を前記追加のディレクトリイメージとする方法、
前記障害の発生時点から遡って所定時間内におけるアクセス頻度が高いファイルを前記

追加のディレクトリイメージとする方法、
のうちの少なくともいずれかである
情報処理システムにおける障害復旧方法。
【請求項６】
請求項２に記載の情報処理システムにおける障害復旧方法であって、
前記第１サーバ装置は、前記第１ファイルシステムに割り当てられている前記第１ストレージ装置の記憶領域の残容量が予め設定された閾値未満になっている場合に、前記第１ストレージ装置に記憶しているファイルのデータのうち、所定の選出基準に従って選出したファイルのデータのメタデータについては前記第１ストレージ装置に残すとともに当該データの実体については前記第１ストレージ装置から削除して前記残容量を確保するスタブ化を行い、
前記第２サーバ装置は、前記第１サーバ装置に送信した、前記ディレクトリイメージ又は前記追加のディレクトリイメージが再びスタブ化される現象である再スタブ化の発生頻度が予め設定された閾値以上になっているか否か、もしくは、再スタブ化の発生時間間隔が予め設定された閾値未満になっているか否かを監視し、
前記第２サーバ装置は、再スタブ化の発生頻度が予め設定された閾値以上になっているか、もしくは、再スタブ化の発生時間間隔が予め設定された閾値未満になっている場合に、前記第１サーバ装置への前記ディレクトリイメージ又は前記追加のディレクトリイメージの送信を抑制する
情報処理システムにおける障害復旧方法。
【請求項７】
請求項６に記載の情報処理システムにおける障害復旧方法であって、
前記抑制は、
データＩ／Ｏ要求がファイルのメタデータのみを対象としている場合に当該ファイルの実体を復元しないようにする方法、
前記所定の選出方法の一つ以上を用いた前記追加のディレクトリイメージの選出を行っている場合に、更に別の選出方法を重複して適用するようにする方法、
前記データＩ／Ｏ要求の対象になっているファイルよりも高い重要度が設定されているファイルのメタデータ及び／又は実体を、前記追加のディレクトリイメージとして選出するようにしている場合に、選出基準としている前記重要度を更に高く設定するようにする方法、
前記障害の発生時点から遡って所定時間内におけるアクセス頻度が前記データＩ／Ｏ要求の対象になっているファイルよりも高いファイルを、前記追加のディレクトリイメージとして選出するようにしている場合に、選出基準としている前記アクセス頻度を更に高く設定するようにする方法、
のうちの少なくともいずれかの方法で行われる
情報処理システムにおける障害復旧方法。
【請求項８】
請求項６に記載の情報処理システムにおける障害復旧方法であって、
前記第１サーバ装置は、前記第１ストレージ装置に記憶しているファイルが現在スタブ化されているか否かを示す情報を前記第２サーバ装置に随時送信し、
前記第２サーバ装置は、前記ディレクトリイメージ又は前記追加のディレクトリイメージの前記第１サーバ装置への送信履歴を管理し、
前記第２サーバ装置は、前記情報及び前記送信履歴に基づき、前記再スタブ化の発生頻度、もしくは、前記再スタブ化の発生時間間隔を取得する
情報処理システムにおける障害復旧方法。
【請求項９】
請求項１に記載の情報処理システムにおける障害復旧方法であって、
前記第１サーバ装置は、前記第１ストレージ装置に記憶されているファイルのデータと、前記第２ストレージ装置に記憶されているファイルのデータと、を一致させる旨を要求

する同期要求が発生すると、前記第１ストレージ装置に記憶されているファイルのデータ又は前回同期時からの更新差分を前記第２サーバ装置に送信し、
前記第２サーバ装置は、前記ファイルのデータ又は前記更新差分に基づき、前記第１ストレージ装置に記憶されているファイルのデータの複製として前記第２ストレージ装置に記憶しているファイルのデータを更新する
情報処理システムにおける障害復旧方法。
【請求項１０】
請求項８に記載の情報処理システムにおける障害復旧方法であって、
前記第１サーバ装置は、前記第１ストレージ装置に記憶されているファイルのデータと、前記第２ストレージ装置に記憶されているファイルのデータとを一致させる旨を要求する同期要求が発生すると、前記第１ストレージ装置に記憶されているファイルのデータ又は前回同期時からの更新差分を前記第２サーバ装置に送信し、
前記第２サーバ装置は、前記ファイルのデータ又は前記更新差分に基づき、前記第１ストレージ装置に記憶されているファイルのデータの複製として前記第２ストレージ装置に記憶しているファイルのデータを更新し、
ファイルが現在スタブ化されているか否かを示す前記情報は、前記第１ストレージ装置又は前記第２ストレージ装置に記憶されているファイルのデータのうちメタデータに含まれている
情報処理システムにおける障害復旧方法。
【請求項１１】
請求項１に記載の情報処理システムにおける障害復旧方法であって、
前記ディレクトリイメージには、ディレクトリの階層構造、ディレクトリのメタデータ、ファイルのメタデータ、ファイルの実体
のうちの少なくともいずれかが含まれている
情報処理システムにおける障害復旧方法。
【請求項１２】
請求項１に記載の情報処理システムにおける障害復旧方法であって、
前記第２サーバ装置は、前記第１サーバ装置から前記要求が送られてきた場合、前記要求の対象である前記ディレクトリイメージを前記第２ストレージ装置から読み出して前記第１サーバ装置に送信するとともに、所定の選出方法に従って選出した、前記ディレクトリイメージとは異なる追加のディレクトリイメージを、前記第２ストレージ装置から読み出して前記第１サーバ装置に送信し、
前記第１サーバ装置は、前記第２サーバ装置から送られてくる前記ディレクトリイメージに基づき前記データＩ／Ｏ要求を処理するとともに、前記ディレクトリイメージ及び前記第２サーバ装置から送られてくる前記追加のディレクトリイメージを、前記第１ストレージ装置に復元し、
前記第２サーバ装置は、前記第１サーバ装置が受け付けた前記データＩ／Ｏ要求が所定の条件を満たす場合に、所定の選出方法に従って前記追加のディレクトリイメージを選出し、
前記所定の条件は、
前記データＩ／Ｏ要求の対象になっているファイルのデータサイズが、現在から遡って所定時間内に発生したデータＩ／Ｏ要求が対象としていたファイルのデータサイズの平均値よりも小さいこと、
前記データＩ／Ｏ要求の対象になっているファイルのデータサイズが予め設定された閾値よりも小さいこと
のうちの少なくともいずれかであり、
前記所定の選出方法は、
前記第１ストレージ装置に既に復元されているディレクトリの配下に存在するファイルのメタデータ及び／又は実体を前記追加のディレクトリイメージとする方法、
前記第１ストレージ装置に既に復元されているディレクトリの配下に存在するディレク

トリのメタデータを前記追加のディレクトリイメージとする方法、
前記障害が発生する前に前記第１ストレージ装置に実体が格納されていたファイルの実体を前記追加のディレクトリイメージとする方法、
重要度が高く設定されているファイルのメタデータ及び／又は実体を前記追加のディレクトリイメージとする方法、
前記障害の発生時点から遡って所定時間内におけるアクセス頻度が高いファイルを前記追加のディレクトリイメージとする方法、
のうちの少なくともいずれかであり、
前記第１サーバ装置は、前記第１ファイルシステムに割り当てられている前記第１ストレージ装置の記憶領域の残容量が予め設定された閾値未満になっている場合に、前記第１ストレージ装置に記憶しているファイルのデータのうち、所定の選出基準に従って選出したファイルのデータのメタデータについては前記第１ストレージ装置に残すとともに当該データの実体については前記第１ストレージ装置から削除して前記残容量を確保するスタブ化を行い、
前記第２サーバ装置は、前記第１サーバ装置に送信した、前記ディレクトリイメージ又は前記追加のディレクトリイメージが再びスタブ化される現象である再スタブ化の発生頻度が予め設定された閾値以上になっているか否か、もしくは、再スタブ化の発生時間間隔が予め設定された閾値未満になっているか否かを監視し、
前記第２サーバ装置は、再スタブ化の発生頻度が予め設定された閾値以上になっているか、もしくは、再スタブ化の発生時間間隔が予め設定された閾値未満になっている場合に、前記第１サーバ装置への前記ディレクトリイメージ又は前記追加のディレクトリイメージの送信を抑制し、
前記抑制は、
データＩ／Ｏ要求がファイルのメタデータのみを対象としている場合に当該ファイルの実体を復元しないようにする方法、
前記所定の選出方法の一つ以上を用いた前記追加のディレクトリイメージの選出を行っている場合に、更に別の選出方法を重複して適用するようにする方法、
前記データＩ／Ｏ要求の対象になっているファイルよりも高い重要度が設定されているファイルのメタデータ及び／又は実体を、前記追加のディレクトリイメージとして選出するようにしている場合に、選出基準としている前記重要度を更に高く設定するようにする方法、
前記障害の発生時点から遡って所定時間内におけるアクセス頻度が前記データＩ／Ｏ要求の対象になっているファイルよりも高いファイルを、前記追加のディレクトリイメージとして選出するようにしている場合に、選出基準としている前記アクセス頻度を更に高く設定するようにする方法、
のうちの少なくともいずれかの方法で行われ、
前記第１サーバ装置は、前記第１ストレージ装置に記憶しているファイルが現在スタブ化されているか否かを示す情報を前記第２サーバ装置に随時送信し、
前記第２サーバ装置は、前記ディレクトリイメージ又は前記追加のディレクトリイメージの前記第１サーバ装置への送信履歴を管理し、
前記第２サーバ装置は、前記情報及び前記送信履歴に基づき、前記再スタブ化の発生頻度、もしくは、前記再スタブ化の発生時間間隔を取得し、
前記第１サーバ装置は、前記第１ストレージ装置に記憶されているファイルのデータと、前記第２ストレージ装置に記憶されているファイルのデータと、を一致させる旨を要求する同期要求が発生すると、前記第１ストレージ装置に記憶されているファイルのデータ又は前回同期時からの更新差分を前記第２サーバ装置に送信し、
前記第２サーバ装置は、前記ファイルのデータ又は前記更新差分に基づき、前記第１ストレージ装置に記憶されているファイルのデータの複製として前記第２ストレージ装置に記憶しているファイルのデータを更新し、
ファイルが現在スタブ化されているか否かを示す前記情報は、前記第１ストレージ装置

又は前記第２ストレージ装置に記憶されているファイルのデータのうちメタデータに含まれており、
前記ディレクトリイメージには、ディレクトリの階層構造、ディレクトリのメタデータ、ファイルのメタデータ、ファイルの実体
のうちの少なくともいずれかが含まれている
情報処理システムにおける障害復旧方法。
【請求項１３】
第１ファイルシステムを有し、データＩ／Ｏ要求を受け付ける第１サーバ装置と、
第２ファイルシステムを有し、前記第１サーバ装置に通信可能に接続された第２サーバ装置と、を備え、
前記第１サーバ装置は、前記データＩ／Ｏ要求の対象であるファイルのデータを第１ストレージ装置に記憶し、
前記第２サーバ装置は、前記データＩ／Ｏ要求の対象であるファイルのデータを第２ストレージ装置に記憶し、
前記第１サーバ装置が、前記第１ストレージ装置に記憶しているファイルのデータを前記第２サーバ装置に送信し、
前記第２サーバ装置が、前記第１サーバ装置から送られてくる前記データを前記第２ストレージ装置に記憶している情報処理システムであって、
前記第２サーバ装置が、障害からの復旧に際して前記第１サーバ装置が前記データＩ／Ｏ要求の受け付けを開始するのに先立ち、前記第２ストレージ装置に記憶しているディレクトリイメージのうち最上位のディレクトリから所定の下位階層までのディレクトリイメージを、前記第１サーバ装置に送信し、
前記第１サーバ装置が、前記第２サーバ装置から送られてくる前記ディレクトリイメージを前記第１ストレージ装置に復元した後、前記データＩ／Ｏ要求の受け付けを再開し、
前記第１サーバ装置が、前記データＩ／Ｏ要求の受け付け再開後において、受け付けた前記データＩ／Ｏ要求を処理するために必要となるディレクトリイメージが前記第１ストレージ装置に復元されていない場合には、前記ディレクトリイメージを前記第２サーバ装置に要求し、
前記第２サーバ装置が、前記第１サーバ装置から送られてくる前記要求に応じて前記ディレクトリイメージを前記第２ストレージ装置から読み出して前記第１サーバ装置に送信し、
前記第１サーバ装置が、前記第２ストレージ装置から送られてくる前記ディレクトリイメージに基づき前記データＩ／Ｏ要求を処理するとともに、前記ディレクトリイメージを前記第１ストレージ装置に復元する
情報処理システム。
【請求項１４】
請求項１３に記載の情報処理システムであって、
前記第２サーバ装置は、前記第１サーバ装置から前記要求が送られてきた場合、前記要求の対象である前記ディレクトリイメージを前記第２ストレージ装置から読み出して前記第１サーバ装置に送信するとともに、所定の選出方法に従って選出した、前記ディレクトリイメージとは異なる追加のディレクトリイメージを、前記第２ストレージ装置から読み出して前記第１サーバ装置に送信し、
前記第１サーバ装置は、前記第２サーバ装置から送られてくる前記ディレクトリイメージに基づき前記データＩ／Ｏ要求を処理するとともに、前記ディレクトリイメージ及び前記第２サーバ装置から送られてくる前記追加のディレクトリイメージを、前記第１ストレージ装置に復元する
情報処理システム。
【請求項１５】
請求項１４に記載の情報処理システムであって、
前記第１サーバ装置は、前記第１ファイルシステムに割り当てられている前記第１スト

レージ装置の記憶領域の残容量が予め設定された閾値未満になっている場合に、前記第１ストレージ装置に記憶しているファイルのデータのうち、所定の選出基準に従って選出したファイルのデータのメタデータについては前記第１ストレージ装置に残すとともに当該データの実体については前記第１ストレージ装置から削除して前記残容量を確保するスタブ化を行い、

前記第２サーバ装置は、前記第１サーバ装置に送信した、前記ディレクトリイメージ又は前記追加のディレクトリイメージが再びスタブ化される現象である再スタブ化の発生頻度が予め設定された閾値以上になっているか否か、もしくは、再スタブ化の発生時間間隔が予め設定された閾値未満になっているか否かを監視し、

前記第２サーバ装置は、再スタブ化の発生頻度が予め設定された閾値以上になっているか、もしくは、再スタブ化の発生時間間隔が予め設定された閾値未満になっている場合に、前記第１サーバ装置への前記ディレクトリイメージ又は前記追加のディレクトリイメージの送信を抑制する

情報処理システム。

【発明の詳細な説明】

【技術分野】

【０００１】

この発明は、情報処理システムにおける障害復旧方法、及び情報処理システムに関する。

【背景技術】

【０００２】

特許文献１には、階層記憶装置の復旧に要する時間を低減し、階層記憶装置の復旧を高速に行うべく、オペレーティングシステム上で稼動する階層記憶装置において、ファイルの属性情報を含むｉノードを有し且つそのｉノード番号で当該ファイルを一意に識別するファイルシステムが構築された第１の記憶装置と、ファイルシステムのバックアップデータを含むデータを格納する第２の記憶装置とを有する階層記憶装置の復旧方法において、第２の記憶装置上のバックアップデータから第１の記憶装置上にファイルシステムが復旧されるときに、バックアップデータに含まれるｉノード番号を用いて、復旧対象ファイルのｉノード番号を指定し、指定されたｉノード番号をファイルシステムの復旧対象ファイルに割り当てることが記載されている。

【０００３】

特許文献２には、ＨＳＭにおける名前空間のバックアップの世代管理を効率的に行うべく、一次ストレージと二次ストレージを有するＨＳＭの制御を行うＨＳＭ制御方法において、ＨＳＭのバックアップ毎に、該バックアップの世代番号を含む情報である世代情報を作成し、ＨＳＭにおけるファイル毎の名前空間に関する情報である名前空間情報と、世代情報作成ステップにより作成された世代番号を用いて該名前空間に関する情報が有効である世代番号の範囲を示す有効世代番号範囲とを名前空間情報履歴として管理することが記載されている。

【先行技術文献】

【特許文献】

【０００４】

【特許文献１】特開２００５－３１６７０８号公報

【特許文献２】特開２００８－０４０６９９号公報

【発明の概要】

【発明が解決しようとする課題】

【０００５】

ところで、企業における支店や事業所等に設けられている情報処理装置のデータのバックアップをデータセンタ等に設けられているバックアップ装置で管理している情報処理システムにおいて、情報処理装置側に障害が発生した場合には、その復旧に際してバックアップ装置側のバックアップデータの全てを情報処理装置に復元した後に情報処理装置のサ

ービスを再開するため、例えば、バックアップデータのサイズが大きい場合はサービスが再開されるまでに長時間を要し、ユーザの業務等への影響が生じる場合もあった。

【０００６】

本発明はこのような背景に鑑みてなされたもので、障害復旧に際してサービスを迅速に再開することが可能な、情報処理システムにおける障害復旧方法、及び情報処理システムを提供することを主たる目的とする。

【課題を解決するための手段】

【０００７】

上記目的を達成するための本発明の一つは、第１ファイルシステムを有し、データＩ／Ｏ要求を受け付ける第１サーバ装置と、第２ファイルシステムを有し、前記第１サーバ装置に通信可能に接続された第２サーバ装置と、を備え、前記第１サーバ装置は、前記データＩ／Ｏ要求の対象であるファイルのデータを第１ストレージ装置に記憶し、前記第２サーバ装置は、前記データＩ／Ｏ要求の対象であるファイルのデータを第２ストレージ装置に記憶し、前記第１サーバ装置は、前記第１ストレージ装置に記憶しているファイルのデータを前記第２サーバ装置に送信し、前記第２サーバ装置は、前記第１サーバ装置から送られてくる前記データを、前記第１ファイルシステムにおけるディレクトリイメージを保持しつつ前記第２ストレージ装置に記憶している情報処理システムにおける障害の復旧方法であって、前記第２サーバ装置が、障害からの復旧に際し、前記第１サーバ装置が前記データＩ／Ｏ要求の受け付けを開始するのに先立ち、前記第２ストレージ装置に記憶しているディレクトリイメージのうち最上位のディレクトリから所定の下位階層までのディレクトリイメージを前記第１サーバ装置に送信し、前記第１サーバ装置が、前記第２サーバ装置から送られてくる前記ディレクトリイメージを前記第１ストレージ装置に復元した後、前記データＩ／Ｏ要求の受け付けを再開し、前記第１サーバ装置は、前記データＩ／Ｏ要求の受け付け再開後、受け付けた前記データＩ／Ｏ要求を処理するために必要となるディレクトリイメージが前記第１ストレージ装置に復元されていない場合に前記ディレクトリイメージを前記第２サーバ装置に要求し、前記第２サーバ装置は、前記第１サーバ装置から前記要求が送られてくると前記ディレクトリイメージを前記第２ストレージ装置から読み出して送信し、前記第１サーバ装置は、前記第２ストレージ装置から送られてくる前記ディレクトリイメージに基づき前記データＩ／Ｏ要求を処理するとともに、前記ディレクトリイメージを前記第１ストレージ装置に復元することとする。

【０００８】

その他本願が開示する課題やその解決方法については、発明の実施形態の欄及び図面により明らかにされる。

【発明の効果】

【０００９】

本発明によれば、障害復旧に際してサービスを迅速に再開することができる。

【図面の簡単な説明】

【００１０】

【図１】情報処理システム１の概略的な構成を示す図である。

【図２】クライアント装置２のハードウエアの一例である。

【図３】第１サーバ装置３ａ又は第２サーバ装置３ｂとして利用可能な情報処理装置のハードウエアの一例である。

【図４】第１ストレージ装置１０ａ又は第２ストレージ装置１０ｂのハードウエアの一例である。

【図５】チャネル基板１１のハードウエアの一例である。

【図６】プロセッサ基板１２のハードウエアの一例である。

【図７】ドライブ基板１３のハードウエアの一例である。

【図８】ストレージ装置１０が備える基本的な機能を示す図である。

【図９】書き込み処理Ｓ９００を説明するフローチャートである。

【図１０】読み出し処理Ｓ１０００を説明するフローチャートである。

【図１１】クライアント装置２が備える主な機能を示す図である。
【図１２】第１サーバ装置３ａが備える主な機能、及び第１サーバ装置３ａにおいて管理される主な情報（データ）を示す図である。
【図１３】レプリケーション情報管理テーブル３３１の一例である。
【図１４】ファイルアクセスログ３３５の一例である。
【図１５】第２サーバ装置３ｂが備える主な機能、及び第２サーバ装置３ｂにおいて管理される主な情報（データ）を示す図である。
【図１６】リストアログ３６５の一例である。
【図１７】抑制フラグ管理テーブル３６６の一例である。
【図１８】リコールログ３６７の一例である。
【図１９】ｉｎｏｄｅを説明する図である。
【図２０】ｉｎｏｄｅの概念を説明する図である。
【図２１】ｉｎｏｄｅの概念を説明する図である。
【図２２】一般的なｉｎｏｄｅ管理テーブル１９１２の一例である。
【図２３】本実施形態のｉｎｏｄｅ管理テーブル１９１２の一例である。
【図２４】レプリケーション開始処理Ｓ２４００を説明する図である。
【図２５】スタブ化候補選出処理Ｓ２５００を説明する図である。
【図２６】スタブ化処理Ｓ２６００を説明する図である。
【図２７】レプリケーションファイル更新処理Ｓ２７００を説明する図である。
【図２８】レプリケーションファイル参照処理Ｓ２８００を説明する図である。
【図２９】同期処理Ｓ２９００を説明する図である。
【図３０】メタデータアクセス処理Ｓ３０００を説明する図である。
【図３１】スタブ化ファイル実体参照処理Ｓ３１００を説明する図である。
【図３２】スタブ化ファイル実体更新処理Ｓ３２００を説明する図である。
【図３３】仮想マシン復旧処理Ｓ３３００を説明する図である。
【図３４】ディレクトリイメージ事前回復処理Ｓ３４００を説明する図である。
【図３５】オンデマンド復元処理Ｓ３５００を説明する図である。
【図３６】第１ストレージ装置１０ａにディレクトリイメージが復元されていく様子を説明する図である。
【図３７】オンデマンド復元処理（復元対象追加有）Ｓ３７００を説明する図である。
【図３８】再スタブ化回避処理Ｓ３８００を説明する図である。
【図３９】レプリケーション開始処理Ｓ２４００の詳細を説明するフローチャートである。
【図４０】スタブ化候補選出処理Ｓ２５００の詳細を説明するフローチャートである。
【図４１】スタブ化処理Ｓ２６００の詳細を説明するフローチャートである。
【図４２】レプリケーションファイル更新処理Ｓ２７００の詳細を説明するフローチャートである。
【図４３】レプリケーションファイル参照処理Ｓ２８００の詳細を説明するフローチャートである。
【図４４】同期処理Ｓ２９００の詳細を説明するフローチャートである。
【図４５】メタデータアクセス処理Ｓ３０００の詳細を説明するフローチャートである。
【図４６】スタブ化ファイル実体参照処理Ｓ３１００の詳細を説明するフローチャートである。
【図４７】スタブ化ファイル実体更新処理Ｓ３２００の詳細を説明するフローチャートである。
【図４８】仮想マシン復旧処理Ｓ３３００及びディレクトリイメージ事前回復処理Ｓ３４００の詳細を説明するフローチャートである。
【図４９】オンデマンド復元処理Ｓ３５００の詳細を説明するフローチャートである。
【図５０】オンデマンド復元処理（復元対象追加有）Ｓ３７００の詳細を説明するフローチャートである。

【図５１】オンデマンド復元処理（復元対象追加有）Ｓ３７００の詳細を説明するフローチャート（図５０の続き）である。
【図５２】再スタブ化回避処理Ｓ３８００の詳細を説明するフローチャートである。
【発明を実施するための形態】
【００１１】
以下、発明を実施するための形態について図面とともに説明する。
【００１２】
図１に実施形態として説明する情報処理システム１の概略的な構成を示している。同図に示すように、本実施形態として例示する情報処理システム１は、商社や電機メーカ等の企業における支点や事業所などのように、ユーザが実際に業務を行う拠点（以下、エッジ５０（Edge）と称する。）に設けられるハードウエアと、データセンタのように、情報処理システム（アプリケーションサーバ／ストレージシステム等）の管理やクラウドサービスの提供などを行う拠点（以下、コア５１（Core）と称する。）に設けられるハードウエアと、を備える。
【００１３】
同図に示すように、エッジ５０には、第１サーバ装置３ａ、第１ストレージ装置１０ａ、及びクライアント装置２が設けられている。またコア５１には、第２サーバ装置３ｂ及び第２ストレージ装置１０ｂが設けられている。
【００１４】
エッジに設けられている第１サーバ装置３ａは、例えば、エッジに設けられているクライアント装置２に対してファイルを単位としたデータの管理機能を提供するファイルシステムを備えたファイルストレージ装置である。またコアに設けられている第２サーバ装置３ｂは、例えば、エッジに設けられている第１ストレージ装置１０ａに対してデータのアーカイブ（書庫）先として機能するアーカイブ装置である。
【００１５】
同図に示すように、クライアント装置２と第１サーバ装置３ａとは、通信ネットワーク５を介して通信可能に接続している。また第１サーバ装置３ａと第１ストレージ装置１０ａとは、第１ストレージネットワーク６ａを介して通信可能に接続している。また第２サーバ装置３ｂと第２ストレージ装置１０ｂとは、第２ストレージネットワーク６ｂを介して通信可能に接続している。また第１サーバ装置３ａと第２サーバ装置３ｂとは、通信ネットワーク７を介して通信可能に接続している。
【００１６】
通信ネットワーク５及び通信ネットワーク７は、例えば、ＬＡＮ（Local Area Network）、ＷＡＮ（Wide Area Network）、インターネット、公衆通信網、専用線などである。第１ストレージネットワーク６ａ及び第２ストレージネットワーク６ｂは、例えば、ＬＡＮ、ＷＡＮ、ＳＡＮ（Storage Area Network）、インターネット、公衆通信網、専用線などである。
【００１７】
通信ネットワーク５、通信ネットワーク７、第１ストレージネットワーク６ａ、及び第２ストレージネットワーク６ｂを介して行われる通信は、例えば、ＴＣＰ／ＩＰ、ｉＳＣＳＩ（internet Small Computer System Interface）、ファイバーチャネルプロトコル（Fibre Channel Protocol）、ＦＩＣＯＮ（Fibre Connection）（登録商標）、ＥＳＣＯＮ（Enterprise System Connection）（登録商標）、ＡＣＯＮＡＲＣ（Advanced Connection Architecture）（登録商標）、ＦＩＢＡＲＣ（Fibre Connection Architecture）（登録商標）などのプロトコルに従って行われる。
【００１８】
クライアント装置２は、第１サーバ装置３ａを介して第１ストレージ装置１０ａが提供する記憶領域を利用する情報処理装置（コンピュータ）であり、例えば、パーソナルコンピュータ、オフィスコンピュータなどである。クライアント装置２では、ファイルシステム、カーネルやドライバなどのソフトウエアモジュールによって実現されるオペレーティ

ングシステム（Operating System）、及びアプリケーションなどが機能している。

【００１９】

図２にクライアント装置２のハードウエアを示している。同図に示すように、クライアント装置２は、ＣＰＵ２１、揮発性または不揮発性のメモリ２２（ＲＡＭまたはＲＯＭ）、記憶装置２３（例えばハードディスクドライブ、半導体記憶装置（ＳＳＤ（Solid State Drive））、キーボードやマウス等の入力装置２４、液晶モニタやプリンタ等の出力装置２５、及びＮＩＣ（Network Interface Card）（以下、ＬＡＮアダプタ２６１と称する。）等の通信インタフェース（通信Ｉ／Ｆ２６と称する。）を備えている。

【００２０】

第１サーバ装置３ａは、第１ストレージ装置１０ａが提供する記憶領域を利用してクライアント装置２に情報処理サービスを提供する情報処理装置である。第１サーバ装置３ａは、パーソナルコンピュータ、メインフレーム（Mainframe）、オフィスコンピュータなどを用いて構成されている。第１サーバ装置３ａは、第１ストレージ装置１０ａが提供する記憶領域へのアクセスに際し、データＩ／Ｏ要求（データ書き込み要求、データ読み出し要求等）を含んだデータフレーム（以下、フレームと略記する。）を、第１ストレージネットワーク６ａを介して第１ストレージ装置１０ａに送信する。尚、上記フレームは、例えば、ファイバーチャネルのフレーム（ＦＣフレーム（FC: Fibre Channel））である。

【００２１】

第２サーバ装置３ｂは、第２ストレージ装置１０ｂが提供する記憶領域を利用して情報処理を行う情報処理装置である。第２サーバ装置３ｂは、パーソナルコンピュータ、メインフレーム、オフィスコンピュータなどを用いて構成されている。第２サーバ装置３ｂは、第２ストレージ装置１０ｂが提供する記憶領域へのアクセスに際し、データＩ／Ｏ要求を含んだフレームを、第２ストレージネットワーク６ｂを介して第２ストレージ装置１０ｂに送信する。

【００２２】

図３に第１サーバ装置３ａのハードウエアを示している。同図に示すように、第１サーバ装置３ａは、ＣＰＵ３１、揮発性または不揮発性のメモリ３２（ＲＡＭまたはＲＯＭ）、記憶装置３３（例えばハードディスクドライブ、半導体記憶装置（ＳＳＤ））、キーボードやマウス等の入力装置３４、液晶モニタやプリンタ等の出力装置３５、ＮＩＣ（以下、ＬＡＮアダプタ３６１と称する。）やＨＢＡ（以下、ＦＣアダプタ３６２と称する。）等の通信インタフェース（通信Ｉ／Ｆ３６と表記する。）、及びタイマ回路やＲＴＣ等を用いて構成される計時装置３７を備えている。尚、コア側に存在する第２サーバ装置３ｂも第１サーバ装置３ａと同一又は類似のハードウエア構成を有する。

【００２３】

図４に第１ストレージ装置１０ａのハードウエアを示している。第１ストレージ装置１０ａは、例えばディスクアレイ装置である。尚、コア側に存在する第２ストレージ装置１０ｂも第１ストレージ装置１０ａと同一又は類似のハードウエア構成を有する。ストレージ装置１０は、サーバ装置３（第１サーバ装置３ａ又は第２サーバ装置３ｂ。以下同様）から送られてくるデータＩ／Ｏ要求を受け付け、受け付けたデータＩ／Ｏ要求に応じて記録媒体にアクセスしてサーバ装置３にデータやレスポンスを送信する。

【００２４】

同図に示すように、ストレージ装置１０は、一つ以上のチャネル基板１１、一つ以上のプロセッサ基板１２（Micro Processor）、一つ以上のドライブ基板１３、キャッシュメモリ１４（Cache Memory）、共有メモリ１５（Shared Memory）、内部スイッチ１６、記憶装置１７、及び保守装置１８（SVP: SerVice Processor）を備える。チャネル基板１１、プロセッサ基板１２、ドライブ基板１３、キャッシュメモリ１４、及び共有メモリ１５は、内部スイッチ１６を介して互いに通信可能に接続されている。

【００２５】

チャネル基板１１は、サーバ装置３から送られてくるフレームを受信し、受信したフレ

ームに含まれているデータＩ／Ｏ要求についての処理の応答（例えば読み出したデータ、読み出し完了報告、書き込み完了報告）を含んだフレームをサーバ装置３に送信する。

【００２６】

プロセッサ基板１２は、チャネル基板１１が受信したフレームに含まれている上記データＩ／Ｏ要求に応じて、チャネル基板１１、ドライブ基板１３、及びキャッシュメモリ１４の間で行われるデータ転送（ＤＭＡ（Direct Memory Access）等を用いた高速大容量のデータ転送)に関する処理を行う。プロセッサ基板１２は、キャッシュメモリ１４を介して行われる、チャネル基板１１とドライブ基板１３との間のデータ（記憶装置１７から読み出したデータ、記憶装置１７に書き込むデータ）の転送（引き渡し）や、キャッシュメモリ１４に格納されるデータのステージング（記憶装置１７からのデータの読み出し）及びデステージング（記憶装置１７への書き出し）等を行う。

【００２７】

キャッシュメモリ１４は、高速アクセスが可能なＲＡＭ（Random Access Memory）を用いて構成されている。キャッシュメモリ１４には、記憶装置１７に書き込まれるデータ（以下、書き込みデータと称する。）や記憶装置１７から読み出されたデータ（以下、読み出しデータと記載する。）等が格納される。共有メモリ１５には、ストレージ装置１０の制御に用いられる様々な情報が格納される。

【００２８】

ドライブ基板１３は、記憶装置１７からのデータの読み出しや記憶装置１７へのデータの書き込みに際して記憶装置１７と通信を行う。内部スイッチ１６は、例えば高速クロスバースイッチ（Cross Bar Switch）を用いて構成される。尚、内部スイッチ１６を介して行われる通信は、例えば、ファイバーチャネル、ｉＳＣＳＩ、ＴＣＰ／ＩＰ等のプロトコルに従って行われる。

【００２９】

記憶装置１７は、複数の記憶ドライブ１７１を備えて構成されている。記憶ドライブ１７１は、例えば、ＳＡＳ（Serial Attached SCSI）、ＳＡＴＡ（Serial ATA）、ＦＣ（Fibre Channel）、ＰＡＴＡ（Parallel ATA）、ＳＣＳＩ等のタイプのハードディスクドライブ、半導体記憶装置（ＳＳＤ）等である。

【００３０】

記憶装置１７は、記憶ドライブ１７１を、例えば、ＲＡＩＤ（Redundant Arrays of Inexpensive （or Independent）Disks）等の方式で制御することによって提供される論理的な記憶領域を単位としてサーバ装置３に記憶装置１７の記憶領域を提供する。この論理的な記憶領域は、例えばＲＡＩＤグループ（パリティグループ（Parity Group））を用いて構成される論理装置（ＬＤＥＶ１７２（LDEV: Logical Device））である。

【００３１】

またストレージ装置１０は、サーバ装置３に対してＬＤＥＶ１７２を用いて構成される論理的な記憶領域（以下、ＬＵ（Logical Unit、Logical Volume、論理ボリューム）と称する。）を提供する。ストレージ装置１０は、ＬＵとＬＤＥＶ１７２との対応（関係）を管理しており、ストレージ装置１０は、この対応に基づきＬＵに対応するＬＤＥＶ１７２の特定、もしくは、ＬＤＥＶ１７２に対応するＬＵの特定を行う。

【００３２】

図５にチャネル基板１１のハードウエア構成を示している。同図に示すように、チャネル基板１１は、サーバ装置３と通信するためのポート（通信ポート）を有する外部通信インタフェース（以下、外部通信Ｉ／Ｆ１１１と表記する。）、プロセッサ１１２（フレーム処理チップ及びフレーム転送チップを含む）、メモリ１１３、プロセッサ基板１２と通信するためのポート（通信ポート）を有する内部通信インタフェース（以下、内部通信Ｉ／Ｆ１１４と表記する。）を備えている。

【００３３】

外部通信Ｉ／Ｆ１１１は、ＮＩＣ（Network Interface Card）やＨＢＡ（Host Bus Adaptor）等を用いて構成されている。プロセッサ１１２は、ＣＰＵ（Central Processing U

nit）、ＭＰＵ（Micro Processing Unit）等を用いて構成されている。メモリ１１３は、ＲＡＭ（Random Access Memory）、ＲＯＭ（Read Only Memory）である。メモリ１１３にはマイクロプログラムが格納されている。プロセッサ１１２が、メモリ１１３から上記マイクロプログラムを読み出して実行することにより、チャネル基板１１が提供する各種の機能が実現される。内部通信Ｉ／Ｆ１１４は、内部スイッチ１６を介してプロセッサ基板１２、ドライブ基板１３、キャッシュメモリ１４、共有メモリ１５と通信する。

【００３４】

図６にプロセッサ基板１２のハードウエア構成を示している。プロセッサ基板１２は、内部通信インタフェース（以下、内部通信Ｉ／Ｆ１２１と表記する。）、プロセッサ１２２、及び共有メモリ１５に比べてプロセッサ１２２からのアクセス性能が高い（高速アクセスが可能な）メモリ１２３（ローカルメモリ）を備えている。メモリ１２３にはマイクロプログラムが格納されている。プロセッサ１２２が、メモリ１２３から上記マイクロプログラムを読み出して実行することにより、プロセッサ基板１２が提供する各種の機能が実現される。

【００３５】

内部通信Ｉ／Ｆ１２１は、内部スイッチ１６を介してチャネル基板１１、ドライブ基板１３、キャッシュメモリ１４、及び共有メモリ１５と通信を行う。プロセッサ１２２は、ＣＰＵ、ＭＰＵ、ＤＭＡ（Direct Memory Access）等を用いて構成されている。メモリ１２３は、ＲＡＭ又はＲＯＭである。プロセッサ１２２は、メモリ１２３及び共有メモリ１５のいずれにもアクセスすることができる。

【００３６】

図７にドライブ基板１３のハードウエア構成を示している。ドライブ基板１３は、内部通信インタフェース（以下、内部通信Ｉ／Ｆ１３１と表記する。）、プロセッサ１３２、メモリ１３３、及びドライブインタフェース（以下、ドライブＩ／Ｆ１３４と表記する。）を備えている。メモリ１３３にはマイクロプログラムが格納されている。プロセッサ１３２が、メモリ１３３から上記マイクロプログラムを読み出して実行することにより、ドライブ基板１３が提供する各種の機能が実現される。内部通信Ｉ／Ｆ１３１は、内部スイッチ１６を介して、チャネル基板１１、プロセッサ基板１２、キャッシュメモリ１４、及び共有メモリ１５と通信する。プロセッサ１３２は、ＣＰＵ、ＭＰＵ等を用いて構成されている。メモリ１３３は、例えば、ＲＡＭ、ＲＯＭである。ドライブＩ／Ｆ１３４は、記憶装置１７との間で通信を行う。

【００３７】

図４に示した保守装置１８は、ストレージ装置１０の各構成要素の制御や状態監視を行う。保守装置１８は、パーソナルコンピュータやオフィスコンピュータ等である。保守装置１８は、内部スイッチ１６又はＬＡＮ等の通信手段を介してチャネル基板１１、プロセッサ基板１２、ドライブ基板１３、キャッシュメモリ１４、共有メモリ１５、及び内部スイッチ１６等のストレージ装置１０の構成要素と随時通信を行い、各構成要素から稼働情報等を取得して管理装置１９に提供する。また保守装置１８は管理装置１９から送られてくる制御情報や稼働情報に基づき、各構成要素の設定や制御、保守（ソフトウエアの導入や更新を含む）を行う。

【００３８】

管理装置１９は、保守装置１８とＬＡＮ等を介して通信可能に接続されるコンピュータである。管理装置１９はストレージ装置１０の制御や監視のためのＧＵＩ（Graphical User Interface）やＣＬＩ（Command Line Interface）等を用いたユーザインタフェースを備える。

【００３９】

図８にストレージ装置１０が備える基本的な機能を示している。同図に示すように、ストレージ装置１０は、Ｉ／Ｏ処理部８１１を備えている。Ｉ／Ｏ処理部８１１は、記憶装置１７への書き込みに関する処理を行うデータ書き込み処理部８１１１、記憶装置１７からのデータの読み出しに関する処理を行うデータ読み出し処理部８１１２を備えている。

【００４０】
尚、Ｉ／Ｏ処理部８１１の機能は、ストレージ装置１０のチャネル基板１１やプロセッサ基板１２、ドライブ基板１３が備えるハードウエアにより、又はプロセッサ１１２，１２２，１３２が、メモリ１１３，１２３，１３３に格納されているマイクロプログラムを読み出して実行することにより実現される。
【００４１】
図９は、ストレージ装置１０（第１ストレージ装置１０ａ又は第２ストレージ装置１０ｂ。以下同様。）がサーバ装置３（第１サーバ装置３ａ又は第２サーバ装置３ｂ）からデータ書き込み要求を含んだフレームを受信した場合に、Ｉ／Ｏ処理部８１のデータ書き込み処理部８１１１によって行われる基本的な処理（以下、書き込み処理Ｓ９００と称する。）を説明するフローチャートである。以下、同図とともに書き込み処理Ｓ９００について説明する。尚、以下の説明において、符号の前に付している「Ｓ」の文字は処理ステップを意味している。
【００４２】
同図に示すように、まずサーバ装置３から送信されてくるデータ書き込み要求のフレームが、ストレージ装置１０のチャネル基板１１によって受信される（Ｓ９１１，Ｓ９１２）。
【００４３】
チャネル基板１１は、サーバ装置３からデータ書き込み要求を含んだフレームを受信すると、その旨をプロセッサ基板１２に通知する（Ｓ９１３）。
【００４４】
プロセッサ基板１２は、チャネル基板１１から上記通知を受信すると（Ｓ９２１）、当該フレームのデータ書き込み要求に基づくドライブ書き込み要求を生成し、書き込みデータをキャッシュメモリ１４に格納し、チャネル基板１１に上記通知の受信通知を応答する（Ｓ９２２）。またプロセッサ基板１２は、生成したドライブ書き込み要求をドライブ基板１３に送信する（Ｓ９２３）。
【００４５】
一方、チャネル基板１１は、プロセッサ基板１２からの上記応答を受信すると、サーバ装置３に完了報告を送信し（Ｓ９１４）、サーバ装置３は、チャネル基板１１から完了報告を受信する（Ｓ９１５）。
【００４６】
ドライブ基板１３は、プロセッサ基板１２からドライブ書き込み要求を受信すると、受信したドライブ書き込み要求を書き込み処理待ちキューに登録する（Ｓ９２４）。
【００４７】
ドライブ基板１３は、書き込み処理待ちキューからドライブ書き込み要求を随時読み出し（Ｓ９２５）、読み出したドライブ書き込み要求に指定されている書き込みデータをキャッシュメモリ１４から読み出して、読み出した書き込みデータを記憶装置（記憶ドライブ１７１）に書き込む（Ｓ９２６）。そしてドライブ基板１３は、ドライブ書き込み要求について書き込みデータの書き込みが完了した旨の報告（完了報告）をプロセッサ基板１２に通知する（Ｓ９２７）。
【００４８】
プロセッサ基板１２は、ドライブ基板１３から送られてきた完了報告を受信する（Ｓ９２８）。
【００４９】
図１０は、ストレージ装置１０が、サーバ装置３からデータ読み出し要求を含んだフレームを受信した場合に、ストレージ装置１０のＩ／Ｏ処理部８１１の読み出し処理部８１１２によって行われるＩ／Ｏ処理（以下、読み出し処理Ｓ１０００と称する。）を説明するフローチャートである。以下、同図とともに読み出し処理Ｓ１０００について説明する。
【００５０】

同図に示すように、まずサーバ装置３から送信されてくるフレームが、ストレージ装置１０のチャネル基板１１によって受信される（Ｓ１０１１，Ｓ１０１２）。

【００５１】

チャネル基板１１は、サーバ装置３からデータ読み出し要求を含んだフレームを受信すると、その旨をプロセッサ基板１２及びドライブ基板１３に通知する（Ｓ１０１３）。

【００５２】

ドライブ基板１３は、チャネル基板１１から上記通知を受信すると（Ｓ１０１４）、当該フレームに含まれているデータ読み出し要求に指定されているデータ（例えばＬＢＡ（Logical Block Address）によって指定される）を、記憶装置（記憶ドライブ１７１）から読み出す（Ｓ１０１５）。尚、キャッシュメモリ１４に読み出しデータが存在する場合（キャッシュヒットした場合）には、記憶装置１７からの読み出し処理（Ｓ１０１５）は省略される。

【００５３】

プロセッサ基板１２は、ドライブ基板１３によって読み出されたデータをキャッシュメモリ１４に書き込む（Ｓ１０１６）。そしてプロセッサ基板１２は、キャッシュメモリ１４に書き込んだデータをチャネル基板１１に随時転送する（Ｓ１０１７）。

【００５４】

チャネル基板１１は、プロセッサ基板１２から随時送られてくる読み出しデータを受信すると、それらをサーバ装置３に順次送信する（Ｓ１０１８）。読み出しデータの送信が完了すると、チャネル基板１１は、サーバ装置３に完了報告を送信する（Ｓ１０１９）。サーバ装置３は、読み出しデータ及び完了報告を受信する（Ｓ１０２０，Ｓ１０２１）。

【００５５】

図１１にクライアント装置２が備える主な機能を示している。同図に示すように、クライアント装置２は、アプリケーション２１１、ファイルシステム２１２、及びカーネル／ドライバ２１３の各機能を備える。尚、これらの機能は、クライアント装置２のＣＰＵ２１が、メモリ２２や記憶装置２３に格納されているプログラムを読み出して実行することにより実現される。

【００５６】

ファイルシステム２１２は、クライアント装置２に対してファイル単位又はディレクトリ単位での論理ボリューム（ＬＵ）へのＩ／Ｏ機能を実現する。ファイルシステム２１３は、例えば、ＦＡＴ(File Allocation Table)、ＮＴＦＳ、ＨＦＳ（Hierarchical File System）、ｅｘｔ２（second extended file system）、ｅｘｔ３(third extended file system)、ｅｘｔ４(fourth extended file system)、ＵＤＦ（Universal Disk Format）、ＨＰＦＳ（High Performance File system）、ＪＦＳ（Journaled File System）、ＵＦＳ（Unix File System）、ＶＴＯＣ(Volume Table Of Contents)、ＸＦＳ等である。

【００５７】

カーネル／ドライバ２１３は、オペレーティングシステムのソフトウエアを構成しているカーネルモジュールやドライバモジュールを実行することにより実現される。カーネルモジュールには、クライアント装置２において実行されるソフトウエアについて、プロセスの管理、プロセスのスケジューリング、記憶領域の管理、ハードウエアからの割り込み要求のハンドリング等、オペレーティングシステムが備える基本的な機能を実現するためのプログラムが含まれている。ドライバモジュールには、クライアント装置２を構成しているハードウエアやクライアント装置２に接続して用いられる周辺機器とカーネルモジュールとが通信するためのプログラム等が含まれている。

【００５８】

図１２に第１サーバ装置３ａが備える主な機能、及び第１サーバ装置３ａにおいて管理される主な情報（データ）を示している。同図に示すように、第１サーバ装置３ａでは、仮想環境を提供する仮想化制御部３０５、及びこの仮想化制御部３０５の制御の下で動作する１つ以上の仮想マシン３１０が実現されている。

【００５９】

各仮想マシン３１０では、ファイル共有処理部３１１、ファイルシステム３１２、データ操作要求受付部３１３、データ複製／移動処理部３１４、ファイルアクセスログ取得部３１７、及びカーネル／ドライバ３１８の各機能が実現されている。

【００６０】

尚、仮想環境は、第１サーバ装置３ａのハードウエアと仮想化制御部３０５との間にオペレーティングシステムを介在させるいわゆるホストＯＳ型、もしくは第１サーバ装置３ａのハードウエアと仮想化制御部３０５との間にオペレーティングシステムを介在させないハイパーバイザー型のいずれの方式で実現されていてもよい。またデータ操作要求受付部３１３、データ複製／移動処理部３１４、ファイルアクセスログ取得部３１７の各機能は、ファイルシステム３１２の機能として実現してもよいし、ファイルシステム３１２とは独立した機能として実現してもよい。

【００６１】

同図に示すように、各仮想マシン３１０は、レプリケーション情報管理テーブル３３１、ファイルアクセスログ３３５などの情報（データ）を管理している。これらの情報は、第１ストレージ１０ａから第１サーバ装置３ａに随時読み出されて第１サーバ装置３ａのメモリ３２や記憶装置３３に格納される。

【００６２】

同図に示す機能のうち、ファイル共有処理部３１１は、クライアント装置２にファイルの共有環境を提供する。ファイル共有処理部３１１は、例えば、ＮＦＳ（Network File System）、ＣＩＦＳ（Common Internet File System）、ＡＦＳ（Andrew File System）等のプロトコルに従った機能を提供する。

【００６３】

ファイルシステム３１２は、クライアント装置２に対して、第１ストレージ装置１０ａによって提供される論理ボリューム（ＬＵ）に管理されるファイル（又はディレクトリ）に対するＩ／Ｏ機能を提供する。ファイルシステム３１２は、例えばＦＡＴ(File Allocation Table)、ＮＴＦＳ、ＨＦＳ（Hierarchical File System）、ｅｘｔ２（second extended file system）、ｅｘｔ３(third extended file system)、ｅｘｔ４(fourth extended file system)、ＵＤＦ（Universal Disk Format）、ＨＰＦＳ（High Performance File system）、ＪＦＳ（Journaled File System）、ＵＦＳ（Unix File System）、ＶＴＯＣ(Volume Table Of Contents)、ＸＦＳ等である。

【００６４】

データ操作要求受付部３１３は、クライアント装置２から送信されてくるデータの操作に関する要求（以下、データ操作要求と称する。）を受け付ける。データ操作要求には、後述する、レプリケーション開始要求、レプリケーションファイルに対する更新要求、レプリケーションファイルに対する参照要求、同期要求、メタデータに対するアクセス要求、ファイルの実体に対する参照要求、リコール要求、スタブ化ファイルの実体に対する更新要求などがある。

【００６５】

尚、スタブ化とは、ファイル（又はディレクトリ）のデータのメタデータについては第１ストレージ装置１０ａにて保持するが、ファイル（又はディレクトリ）のデータの実体については第１ストレージ装置１０ａ側では管理せず、第２ストレージ装置１０ｂにおいてのみ保持するようにすることをいう。スタブ化されたファイル（又はディレクトリ）に対してそのファイル（又はディレクトリ）の実体が必要になるようなデータＩ／Ｏ要求を第１サーバ装置３ａが受け付けた場合には、そのファイル（又はディレクトリ）の実体が第２ストレージ装置１０ｂから第１ストレージ装置１０ａに送信される（書き戻される（以下、リコールと称する。））。

【００６６】

データ複製／移動処理部３１４は、後述する、レプリケーション開始処理Ｓ２４００、スタブ化候補選出処理Ｓ２５００、同期処理Ｓ２９００、スタブ化ファイル実体参照処理Ｓ３１００、スタブ化ファイル実体更新処理Ｓ３２００、仮想マシン復旧処理Ｓ３３００

、ディレクトリイメージ事前回復処理Ｓ３４００、オンデマンド復元処理Ｓ３５００、オンデマンド復元処理（復元対象追加有）Ｓ３７００、再スタブ化回避処理Ｓ３８００などにおいて、第１サーバ装置３ａと第２サーバ装置３ｂとの間、もしくは、第１ストレージ装置１０ａと第２ストレージ装置１０ｂとの間での制御情報（フラグやテーブルを含む）の授受やデータ(ファイルのメタデータや実体を含む）の転送、レプリケーション情報管理テーブル３３１、メタデータ３３２などの各種テーブルの管理を行う。

【００６７】

図１２に示すカーネル／ドライバ３１８は、オペレーティングシステムのソフトウエアを構成しているカーネルモジュールやドライバモジュールを実行することにより実現される。カーネルモジュールには、第１サーバ装置３ａにおいて実行されるソフトウエアについて、プロセスの管理、プロセスのスケジューリング、記憶領域の管理、ハードウエアからの割り込み要求のハンドリング等、オペレーティングシステムが備える基本的な機能を実現するためのプログラムが含まれる。ドライバモジュールには、第１サーバ装置３ａを構成しているハードウエアや第１サーバ装置３ａに接続して用いられる周辺機器とカーネルモジュールとが通信するためのプログラムが含まれる。

【００６８】

図１２に示すファイルアクセスログ取得部３１７は、ストレージ装置１０の論理ボリューム（ＬＵ）に格納されているファイルへのアクセス（ファイルの更新（Write, Update）、ファイルの読み出し（Read）、ファイルのオープン（Open）、ファイルのクローズ（Close）等）が行われると、そのアクセスの内容（履歴）を示す情報（以下、アクセスログと称する。）を、計時装置３７から取得される日時情報に基づくタイムスタンプを付与してファイルアクセスログ３３５として記憶する。

【００６９】

図１３にレプリケーション情報管理テーブル３３１の一例を示している。同図に示すように、レプリケーション情報管理テーブル３３１には、レプリケーション先となるホスト名３３１１（例えばＩＰアドレス等のネットワークアドレス。）、スタブ化するか否かの判断に用いる閾値３３１２（後述するスタブ化閾値）が設定される。

【００７０】

図１４にファイルアクセスログ３３５の一例を示している。同図に示すように、ファイルアクセスログ３３５には、アクセス日時３３５１、ファイル名３３５２、及びユーザＩＤ３３５３の各項目からなる一つ以上のレコードで構成されるアクセスログが記録される。

【００７１】

このうちアクセス日時３３５１には、そのファイル（又はディレクトリ）に対するアクセスがされた日時が設定される。ファイル名３３５２には、アクセス対象となったファイル（又はディレクトリ）のファイル名（又はディレクトリ名）が設定される。ユーザＩＤ３３５３には、そのファイル（又はディレクトリ）にアクセスしたユーザのユーザＩＤが設定される。

【００７２】

図１５に第２サーバ装置３ｂが備える主な機能、及び第２サーバ装置３ｂにおいて管理される主な情報（データ）を示している。同図に示すように、第２サーバ装置３ｂは、ファイル共有処理部３５１、ファイルシステム３５２、データ複製／移動処理部３５４、及びカーネル／ドライバ３５８の各機能を備えている。尚、データ複製／移動処理部３５４の機能は、ファイルシステム３５２の機能として実現してもよいし、ファイルシステム３５２とは独立した機能として実現してもよい。

【００７３】

また同図に示すように、第２サーバ装置３ｂは、リストアログ３６５、抑制フラグ管理テーブル３６６、リコールログ３６７、及びファイルアクセスログ３６８を管理している。

【００７４】

ファイル共有処理部３５１は、第１サーバ装置３ａに対してファイルの共有環境を提供する。ファイル共有処理部３５１は、例えばＮＦＳ、ＣＩＦＳ、ＡＦＳ等のプロトコルを用いて実現されている。

【００７５】

ファイルシステム３５２は、第２ストレージ装置１０ｂによって提供される論理ボリューム（ＬＵ）を用い、第１サーバ装置３ａに対してファイル単位又はディレクトリ単位での論理ボリューム（ＬＵ）へのＩ／Ｏ機能を提供する。ファイルシステム３５２は、例えばＦＡＴ、ＮＴＦＳ、ＨＦＳ、ｅｘｔ２、ｅｘｔ３、ｅｘｔ４、ＵＤＦ、ＨＰＦＳ、ＪＦＳ、ＵＦＳ、ＶＴＯＣ、ＸＦＳなどである。

【００７６】

データ複製／移動処理部３５４は、第１ストレージ装置１０ａと第２ストレージ装置１０ｂとの間でのデータの移動や複製に関する処理を行う。

【００７７】

カーネル／ドライバ３５８は、オペレーティングシステムのソフトウエアを構成しているカーネルモジュールやドライバモジュールを実行することにより実現される。カーネルモジュールには、第２サーバ装置３ｂにおいて実行されるソフトウエアについて、プロセスの管理、プロセスのスケジューリング、記憶領域の管理、ハードウエアからの割り込み要求のハンドリング等、オペレーティングシステムが備える基本的な機能を実現するためのプログラムが含まれる。ドライバモジュールには、第２サーバ装置３ｂを構成しているハードウエアや第２サーバ装置３ｂに接続して用いられる周辺機器とカーネルモジュールとが通信するためのプログラムが含まれる。

【００７８】

図１６にリストアログ３６５の一例を示している。リストアログ３６５は、後述するディレクトリイメージの復元（リストア）が行われた場合に、第１サーバ装置３ａ又は第２サーバ装置３ｂによって復元に関する処理の内容が記録される。同図に示すように、リストアログ３６５は、日時３６５１、イベント３６５２、及びリストア対象ファイル３６５３の各項目を有する一つ以上のレコードで構成される。

【００７９】

このうち日時３６５１には、リストアに関するイベントが実行された日時が設定される。イベント３６５２には、実行されたイベントの内容を示す情報（リストア開始、リストア実行等）が設定される。リストア対象ファイル３６５３には、リストアの対象となったファイル（又はディレクトリ）を特定する情報（パス名、ファイル名（又はディレクトリ名）等）が設定される。

【００８０】

図１７に抑制フラグ管理テーブル３６６の一例を示している。抑制フラグ管理テーブル３６６の内容は第２サーバ装置３ｂによって管理される。同図に示すように、抑制フラグ管理テーブル３６６には、後述する再スタブ化回避処理Ｓ３８００において用いられる抑制フラグ３６６１、及び抑制フラグ３６６１の最終更新日時３６６２が管理されている。

【００８１】

図１８にリコールログ３６７の一例を示している。リコールログ３６７の内容は第２サーバ装置３ｂによって生成される。リコールログ３６７には、第２サーバ装置３ｂが第１サーバ装置３ａから受け付けたリコール要求の履歴が管理される。同図に示すように、リコールログ３６７は、日時３６７１及びリコール対象ファイル３６７２の各項目を有する一つ以上のレコードで構成されている。日時３６７１には、リコール要求を受け付けた日時が設定される。リコール対象ファイル３６７２には、受け付けたリコール要求に指定されているリコール対象となるファイル（又はディレクトリ）を特定する情報（パス名、ファイル名等）が設定される。

【００８２】

第２サーバ装置３ｂが管理しているファイルアクセスログ３６８の内容は、第１サーバ装置３ａにおけるファイルアクセスログ３３５の内容に基本的に一致している。第１サー

バ装置３ａからファイルアクセスログ３３５の内容を第２サーバ装置３ｂに随時通知することにより、両者の同一性が確保されている。

【００８３】

次に第１サーバ装置３ａが備えるファイルシステム３１２（第２サーバ３ｂが備えるファイルシステム３５２も同様である。）について詳細に説明する。

【００８４】

図１９は、ファイルシステム３１２が論理ボリューム（ＬＵ）上に管理するデータの構造（以下、ファイルシステム構造１９００と称する。）の一例である。同図に示すように、ファイルシステム構造１９００には、スーパーブロック１９１１、ｉｎｏｄｅ管理テーブル１９１２、ファイルの実体（データ）が格納されるデータブロック１９１３の各記憶領域が含まれている。

【００８５】

このうちスーパーブロック１９１１には、ファイルシステム３１２に関する情報（ファイルシステムが取り扱う記憶領域の容量、使用量、空き容量等）が格納される。スーパーブロック１９１１は、原則としてディスク区画（論理ボリューム（ＬＵ）上に設定されるパーティション）ごとに設けられる。スーパーブロック１９１１に格納される上記情報の具体例として、区画内のデータブロック数、ブロックサイズ、空きブロック数、空きｉｎｏｄｅ数、当該区画のマウント数、最新の整合性チェック時からの経過時間等がある。

【００８６】

ｉｎｏｄｅ管理テーブル１９１２には、論理ボリューム（ＬＵ）に格納されるファイル（又はディレクトリ）の管理情報（以下、ｉｎｏｄｅと称する。）が格納される。ファイルシステム３１２は、１つのファイル（又はディレクトリ）に対して１つのｉｎｏｄｅを対応させて管理している。ｉｎｏｄｅのうちディレクトリに関する情報のみを含むものはディレクトリエントリと称される。ファイルに対するアクセスが行われる場合にはディレクトリエントリを参照してアクセス対象のファイルのデータブロックにアクセスする。例えば、「/home/user-01/a.txt」というファイルにアクセスする場合には、図２０に示すように、ｉｎｏｄｅ番号２→１０→１５→１００とディレクトリエントリを順に辿ってアクセス対象ファイルのデータブロックにアクセスする。

【００８７】

図２１に一般的なファイルシステム（例えば、ＵＮＩＸ（登録商標）系のオペレーティングシステムが備えるファイルシステム）におけるｉｎｏｄｅの概念を示している。また図２２にｉｎｏｄｅ管理テーブル１９１２の一例を示している。

【００８８】

これらの図に示すように、ｉｎｏｄｅには、個々のｉｎｏｄｅを区別する識別子であるｉｎｏｄｅ番号２２１１、当該ファイル（又はディレクトリ）の所有者２２１２、当該ファイル（又はディレクトリ）について設定されているアクセス権２２１３、当該ファイル（又はディレクトリ）のファイルサイズ２２１４、当該ファイル（又はディレクトリ）の最終更新日時２２１５、当該ｉｎｏｄｅがディレクトリエントリである場合に設定される当該ディレクトリの親ディレクトリ２２１６、当該ｉｎｏｄｅがディレクトリエントリである場合に設定される当該ディレクトリの子ディレクトリ２２１７、当該ファイルのデータの実体が格納されるデータブロックを特定する情報（以下、ブロックアドレス２２１８と称する。）などの情報が含まれる。

【００８９】

図２３に示すように、本実施形態のファイルシステム３１２は、図２２に示した通常の一般的なファイルシステムにおけるｉｎｏｄｅ管理テーブル１９１２の内容に加えて、更にスタブ化フラグ２３１１、メタデータ同期要フラグ２３１２、実体同期要フラグ２３１３、レプリケーションフラグ２３１４、リンク先２３１５、及び重要度２３１６を管理している。

【００９０】

尚、レプリケーションによる管理方式やスタブ化による管理方式により、第１ストレー

ジ装置１０ａに記憶されているファイルのメタデータ（図２３に示した各種のフラグを含む。）の複製が第２ストレージ装置１０ｂにも記憶されている場合（レプリケーションされている場合）には、後述する同期処理Ｓ２９００によって一方の装置のメタデータが更新されるとその旨が他方の装置にも通知される。これにより第１ストレージ装置１０ａのメタデータと第２ストレージ装置１０ｂのメタデータの内容の同一性がほぼリアルタイムに確保される。

【００９１】

同図において、スタブ化フラグ２３１１には、当該ｉｎｏｄｅに対応するファイル（又はディレクトリ）がスタブ化されているか否かを示す情報が設定される。ここでスタブ化とは、第１ストレージ装置１０ａから第２ストレージ装置１０ｂにファイル（又はディレクトリ）を移動（マイグレーション）するに際し、移動元の第１ストレージ装置１０ａからそのファイルのデータのうち実体のみ削除し、そのファイルのデータのうちメタデータについては削除せずに移動元の第１ストレージ装置１０ａに残すことをいう。

【００９２】

尚、スタブ（stub）とは、その場合に第１ストレージ装置１０ａに残留するメタデータのことをいう。当該ｉｎｏｄｅに対応するファイル（又はディレクトリ）がスタブ化されている場合はスタブ化フラグ２３１１にＯＮが設定され、スタブ化されていない場合はスタブ化フラグ２３１１にＯＦＦが設定される。

【００９３】

メタデータ同期要フラグ２３１２には、複製元の第１ストレージ装置１０ａのファイル（又はディレクトリ）のメタデータと複製先の第２ストレージ装置１０ｂのファイル（又はディレクトリ）のメタデータとの間で同期をとる必要（内容を一致させる必要）が有るか否かを示す情報が設定される。メタデータの同期が必要な場合はメタデータ同期要フラグ２３１２にＯＮが設定され、同期が不要な場合はメタデータ同期要フラグ２３１２にＯＦＦが設定される。

【００９４】

実体同期要フラグ２３１３には、複製元の第１ストレージ装置１０ａのファイルのデータの実体と複製先の第２ストレージ装置１０ｂのファイルのデータの実体との間で同期をとる必要（内容を一致させる必要）が有るか否かを示す情報が設定される。ファイルのデータの実体について同期が必要な場合は実体同期要フラグ２３１３にＯＮが設定され、同期が不要な場合は実体同期要フラグ２３１３にＯＦＦが設定される。

【００９５】

メタデータ同期要フラグ２３１２及び実体同期要フラグ２３１３は、後述する同期処理Ｓ２９００において随時参照される。メタデータ同期要フラグ２３１２、もしくは実体同期要フラグ２３１３がＯＮになっている場合は、第１ストレージ装置１０ａのメタデータ又は実体と、その複製である第２ストレージ装置１０ｂのメタデータ又は実体とが自動的に同期される。

【００９６】

レプリケーションフラグ２３１４には、そのｉｎｏｄｅに対応するファイル（又はディレクトリ）が、現在後述するレプリケーション管理方式による管理の対象になっているか否かを示す情報が設定される。当該ｉｎｏｄｅに対応するファイルが、現在、レプリケーション管理方式による管理の対象になっている場合はレプリケーションフラグ２３１４にＯＮが設定され、レプリケーションによる管理の対象になっていない場合はレプリケーションフラグ２３１４にＯＦＦが設定される。

【００９７】

リンク先２３１５には、そのｉｎｏｄｅに対応するファイルが、後述するレプリケーション管理方式で管理されている場合は、そのファイルの複製先を示す情報（例えば、格納先を特定するパス名、ＲＡＩＤグループの識別子、ブロックアドレス、ＵＲＬ（Uniform Resource Locator）、ＬＵ等）が設定される。

【００９８】

重要度２３１６には、ファイルの重要度が設定される。重要度２３１６の内容は、例えばユーザがクライアント装置２において設定する。また重要度２３１６は負荷分散等を目的として設定される場合もある。

【００９９】

＝概略的な動作＝

次に、以上の構成からなる情報処理システム１の動作について説明する。

【０１００】

図２４は、第１サーバ装置３ａが、１ストレージ装置１０ａに格納されているファイルを対象としたレプリケーションを開始させる旨の要求（以下、レプリケーション開始要求と称する。）を受け付けた場合に、情報処理システム１において行われる処理（以下、レプリケーション開始処理Ｓ２４００と称する）を説明する図である。

【０１０１】

第１サーバ装置３ａは、クライアント装置２からレプリケーション開始要求を受け付けると、当該要求に対象として指定されているファイルについてレプリケーションによる管理方式による管理を開始する。尚、第１サーバ装置３ａは、通信ネットワーク５を介してクライアント装置２からレプリケーション開始要求を受け付ける以外に、例えば、当該第１サーバ装置３ａにおいて内部的に発生するレプリケーション開始要求も受け付ける。

【０１０２】

ここでレプリケーションによる管理方式とは、ファイルのデータ（メタデータ及び実体）を、第１ストレージ装置１０ａ及び第２ストレージ装置１０ｂの双方において管理する方式である。

【０１０３】

レプリケーションによる管理方式では、第１ストレージ装置１０ａに格納されているファイルの実体又はメタデータが更新されると、このファイルの複製(又はアーカイブファイル）として管理されている、第２ストレージ装置１０ｂ側のファイルのメタデータ又は実体が、同期又は非同期に更新される。レプリケーションによる管理方式が行われることにより、第１ストレージ装置１０ａに格納されているファイルのデータ（メタデータ又は実体）と、その複製として第２ストレージ装置１０ｂに格納されているファイルのデータ（メタデータ又は実体）の一致性が、同期又は非同期に確保（保証）される。

【０１０４】

尚、第２ストレージ装置１０ｂ側のファイル（アーカイブファイル）におけるメタデータはファイルの実体として管理してもよい。そのようにすることで、第１サーバ装置３ａのファイルシステム３１２と第２サーバ装置３ｂのファイルシステム３５２の仕様が相違する場合でもレプリケーションによる管理方式を実現することができる。

【０１０５】

同図に示すように、レプリケーション開始要求を受け付けると（Ｓ２４１１）、第１サーバ装置３ａは、受け付けたレプリケーション開始要求に指定されているファイルのデータ（メタデータ及び実体）を第１ストレージ装置１０ａから読み出し、読み出したファイルのデータを第２サーバ装置３ｂに送信する（Ｓ２４１２）。

【０１０６】

第２サーバ３ｂは、第１サーバ装置３ａから送られてくる、上記ファイルのデータを受信すると、受信したデータを第２ストレージ装置１０ｂに格納する（Ｓ２４１３）。

【０１０７】

尚、上記転送に際し、第１サーバ装置３ａのデータ複製／移動処理部３１４は、転送元ファイルのレプリケーションフラグ２３１４をＯＮに設定する（Ｓ２４１４）。

【０１０８】

図２５は、第１ストレージ装置１０ａに格納されている、レプリケーション管理方式により管理されているファイル（レプリケーションフラグ２３１４がＯＮに設定されているファイル。以下、レプリケーションファイルと称する。）を前述したスタブ化の候補として設定する際に、情報処理システム１において行われる処理（以下、スタブ化候補選出処

理Ｓ２５００と称する。）を説明する図である。以下、同図とともにスタブ化候補選出処理Ｓ２５００について説明する。

【０１０９】

第１サーバ装置３ａは、ファイル格納領域の残容量を随時（リアルタイム、定期的、予め設定されたタイミング等）監視している。

【０１１０】

第１サーバ装置３ａは、ファイルシステム３１２に対してファイルの格納領域として割り当てられている第１ストレージ装置１０ａの記憶領域（以下、ファイル格納領域と称する。）の残容量が、予め設定された閾値（以下、スタブ化閾値と称する。）未満になると、所定の選出基準に従い、第１ストレージ装置１０ａに格納されているレプリケーションファイルの中からスタブ化の候補を選出する（Ｓ２５１１）。尚、上記所定の選出基準としては、例えば、最終更新日時が古い順、アクセス頻度の低い順などがある。

【０１１１】

次に第１サーバ装置３ａは、スタブ化の候補を選出すると、選出したレプリケーションファイルのスタブ化フラグ２３１１をＯＮ、レプリケーションフラグ２３１４をＯＦＦ、メタデータ同期要フラグ２３１２をＯＮに夫々設定する（Ｓ２５１２）。尚、第１サーバ装置３ａは、ファイル格納領域の残容量を、例えば、ファイルシステム３１２が管理している情報から取得する。

【０１１２】

図２６は、スタブ化候補選出処理Ｓ２５００によってスタブ化候補として選出されたファイルを実際にスタブ化する際、情報処理システム１において行われる処理（以下、スタブ化処理Ｓ２６００と称する。）を説明する図である。スタブ化処理Ｓ２６００は、例えば、予め設定されたタイミング（例えばスタブ化候補選出処理Ｓ２５００が行われたのに続いて）で行われる。以下、同図とともにスタブ化処理Ｓ２６００について説明する。

【０１１３】

同図に示すように、第１サーバ装置３ａは、第１ストレージ装置１０ａのファイル格納領域に格納されているファイルの中から、スタブ化候補として選出されているファイル（スタブ化フラグ２３１１がＯＮに設定されているファイル）を一つ以上抽出する（Ｓ２６１１）。

【０１１４】

そして第１サーバ装置３ａは、抽出したファイルの実体を第１ストレージ装置１０ａから削除するとともに、抽出したファイルのメタデータからそのファイルの第１ストレージ装置１０ａの格納先を示す情報に無効な値を設定し（例えば、メタデータの当該ファイルの格納先を設定する欄（例えばブロックアドレス２２１８を設定する欄）にＮＵＬＬ値やゼロを設定する。）、スタブ化候補として選出されているファイルを実際にスタブ化する。またこのとき第１サーバ装置３ａは、メタデータ同期要フラグ２３１２をＯＮに設定する（Ｓ２６１２）。

【０１１５】

図２７は、第１サーバ装置３ａが、クライアント装置２から第１ストレージ装置１０ａのファイル格納領域に格納されているレプリケーションファイルに対する更新要求を受け付けた場合に、情報処理システム１において行われる処理（以下、レプリケーションファイル更新処理Ｓ２７００と称する。）を説明する図である。以下、同図とともにレプリケーションファイル更新処理Ｓ２７００について説明する。

【０１１６】

第１サーバ装置３ａは、レプリケーションファイルに対する更新要求を受け付けると（Ｓ２７１１）、その第１ストレージ装置１０ａに格納されているそのレプリケーションファイルのデータ（メタデータ、実体）を、受け付けた更新要求に従って更新する（Ｓ２７１２）。

【０１１７】

そして第１サーバ装置３ａは、メタデータを更新した場合はそのレプリケーションファ

イルのメタデータ同期要フラグ２３１２をＯＮに設定し、レプリケーションファイルの実体を更新した場合はそのレプリケーションファイルの実体同期要フラグ２３１３をＯＮに設定する（Ｓ２７１３）。

【０１１８】

図２８は、第１サーバ装置３ａのファイルシステム３１２が、クライアント装置２から、第１ストレージ装置１０ａのファイル格納領域に格納されているレプリケーションファイルに対する参照要求を受け付けた場合に情報処理システム１において行われる処理（以下、レプリケーションファイル参照処理Ｓ２８００と称する。）を説明する図である。以下、同図とともにレプリケーションファイル参照処理Ｓ２８００について説明する。

【０１１９】

第１サーバ装置３ａのファイルシステム３１２は、レプリケーションファイルに対する更新要求を受け付けると（Ｓ２８１１）、そのレプリケーションファイルのデータ（メタデータ又は実体）を第１ストレージ装置１０ａから読み出し（Ｓ２８１２）、読み出したデータに基づきクライアント装置２に対して応答する情報を生成し、生成した応答情報をクライアント装置２に送信する（Ｓ２８１３）。

【０１２０】

図２９は、第１サーバ装置３ａが、クライアント装置２から、第１ストレージ装置１０ａに格納されているレプリケーションファイルとその第２ストレージ装置１０ｂ側のファイルの内容とを一致させる要求（以下、同期要求と称する。）を受け付けた際に情報処理システム１において行われる処理（以下、同期処理Ｓ２９００と称する。）を説明する図である。以下、同図とともに同期処理Ｓ２９００について説明する。

【０１２１】

尚、同期処理Ｓ２９００は、クライアント装置２からの同期要求を受け付けた場合以外の事象を契機として開始するようにしてもよい。例えば、予め設定されたタイミング（リアルタイム、定期的等）が到来したことを契機として、第１サーバ装置３ａが自発的に同期処理Ｓ２９００を開始するようにしてもよい。

【０１２２】

第１サーバ装置３ａは、クライアント装置２からレプリケーションファイルの同期要求を受け付けると（Ｓ２９１１）、第１ストレージ装置１０ａのファイル格納領域に格納されているレプリケーションファイルのうち、メタデータ同期要フラグ２３１２又は実体同期要フラグ２３１３の少なくともいずれかがＯＮに設定されているファイルを取得する（Ｓ２９１２）。

【０１２３】

そして第１サーバ装置３ａは、取得したファイルのメタデータ又は実体を第２サーバ装置３ｂに送信するとともに、当該レプリケーションファイルのメタデータ同期要フラグ２３１２又は実体同期要フラグ２３１３をＯＦＦに設定する（Ｓ２９１３）。

【０１２４】

第２サーバ装置３ｂは、メタデータ又は実体を受信すると（Ｓ２９１３）、受信したメタデータ又は実体に対応する、第２ストレージ装置１０ｂに格納されているファイルのメタデータ又は実体を、受信したメタデータ又は実体に基づき更新する（Ｓ２９１４）。尚、第１サーバ装置３ａから第２サーバ装置３ｂにメタデータ又は実体の全体を送信するのではなく、前回同期時からの更新差分のみを送信するようにしてもよい。

【０１２５】

以上に説明した同期処理Ｓ２９００が行われることにより、第１ストレージ装置１０ａに格納されているファイルのデータ（メタデータ及び実体）と、当該ファイルに対応する第２ストレージ装置１０ｂに格納されているファイルのデータ（メタデータ及び実体）とが同期される。

【０１２６】

図３０は、第１サーバ装置３ａのファイルシステム３１２が、クライアント装置２等から、スタブ化されているファイル（スタブ化フラグ２３１１がＯＮに設定されているファ

イル）のメタデータに対するアクセス要求（参照要求又は更新要求）を受け付けた場合に情報処理システム１において行われる処理（以下、メタデータアクセス処理Ｓ３０００と称する。）を説明する図である。以下、同図とともにメタデータアクセス処理Ｓ３０００について説明する。

【０１２７】

同図に示すように、第１サーバ装置３ａは、スタブ化されているファイルのメタデータに対するアクセス要求を受け付けると（Ｓ３０１１）、アクセス要求の対象になっている第１ストレージ装置１０ａのメタデータを取得し、アクセス要求の内容に従い参照（読み出したメタデータに基づく応答情報のクライアント装置２への送信）、又はメタデータの更新を行う（Ｓ３０１２）。またメタデータの内容を更新した場合には、そのファイルのメタデータ同期要フラグ２３１２にＯＮを設定する（Ｓ３０１３）。

【０１２８】

このように、第１サーバ装置３ａは、スタブ化されているファイルに対するアクセス要求が発生し、そのアクセス要求がそのファイルのメタデータのみを対象としている場合には、第１ストレージ装置１０ａに格納されているメタデータを用いてアクセス要求を処理する。このため、アクセス要求がそのファイルのメタデータのみを対象としている場合にはクライアント装置２に迅速に応答を返すことができる。

【０１２９】

図３１は、第１サーバ装置３ａが、クライアント装置２から、スタブ化されているファイル（スタブ化フラグ２３１１がＯＮに設定されているファイル。以下、スタブ化ファイルと称する。）の実体に対する参照要求を受け付けた場合に情報処理システム１において行われる処理（以下、スタブ化ファイル実体参照処理Ｓ３１００と称する。）を説明する図である。以下、同図とともにスタブ化ファイル実体参照処理Ｓ３１００について説明する。

【０１３０】

第１サーバ装置３ａは、クライアント装置２からスタブ化ファイルの実体に対する参照要求を受け付けると（Ｓ３１１１）、取得したメタデータを参照して当該スタブ化ファイルの実体が第１ストレージ装置１０ａに格納されているか否かを判断する（Ｓ３１１２）。ここでこの判断は、例えば、取得したメタデータにスタブ化ファイルの実体の格納先を示す情報（例えばブロックアドレス２２１８）に有効な値が設定されているか否かに基づいて行う。

【０１３１】

上記判断の結果、スタブ化ファイルの実体が第１ストレージ装置１０ａに格納されている場合には、第１サーバ装置３ａは、第１ストレージ装置１０ａから当該スタブ化ファイルの実体を読み出し、読み出した実体に基づきクライアント装置２に応答する情報を生成し、生成した応答情報をクライアント装置２に送信する（Ｓ３１１３）。

【０１３２】

一方、上記判断の結果、スタブ化ファイルの実体が第１ストレージ装置１０ａに格納されていない場合には、第１サーバ装置３ａは、第２サーバ装置３ｂに対してスタブ化ファイルの実体の提供を要求する（以下、リコール要求と称する。）（Ｓ３１１４）。 尚、実体の取得要求は、必ずしも一度の取得要求によって実体の全体を取得する要求でなくてもよく、例えば実体の一部のみを複数回要求するようにしてもよい。

【０１３３】

第１サーバ装置３ａは、上記取得要求に応じて第２サーバ装置３ｂから送られてくるスタブ化ファイルの実体を受信すると（Ｓ３１１５）、受信した実体に基づき応答情報を生成し、生成した応答情報をクライアント装置２に送信する（Ｓ３１１６）。

【０１３４】

また第１サーバ装置３ａは、上記第２サーバ装置３ｂから受信した実体を第１ストレージ装置１０ａに格納し、当該スタブ化ファイルのメタデータの当該ファイルの実体の格納先を示す情報（例えばブロックアドレス２２１８）に、当該ファイルの第１ストレージ装

置１０ａにおける格納先を示す内容を設定する。また第１サーバ装置３ａは、当該ファイルのスタブ化フラグ２３１１をＯＦＦに、レプリケーションフラグ２３１４をＯＮに、メタデータ同期要フラグ２３１２をＯＮに、夫々設定する（当該ファイルをスタブ化ファイルからレプリケーションファイルに変更する。）（Ｓ３１１７）。

【０１３５】

尚、メタデータ同期要フラグ２３１２をＯＮに設定するのは、第１ストレージ装置１０ａと第２ストレージ装置１０ｂとの間で、当該スタブ化ファイルのスタブ化フラグ２３１１及びレプリケーションフラグ２３１４の内容を事後的に自動的に同期させるためである。

【０１３６】

図３２は、第１サーバ装置３ａが、クライアント装置２からスタブ化ファイルの実体に対する更新要求を受け付けた場合に、情報処理システム１において行われる処理（以下、スタブ化ファイル実体更新処理Ｓ３２００と称する。）を説明する図である。以下、同図とともにスタブ化ファイル実体更新処理Ｓ３２００について説明する。

【０１３７】

第１サーバ装置３ａは、スタブ化ファイルの実体に対する更新要求を受け付けると（Ｓ３２１１）、更新要求の対象になっているスタブ化ファイルのメタデータを取得し、取得したメタデータに基づき当該スタブ化ファイルの実体が第１ストレージ装置１０ａに格納されているか否かを判断する（Ｓ３２１２）。尚、判断方法はスタブ化ファイル実体参照処理Ｓ３１００の場合と同様である。

【０１３８】

上記判断の結果、スタブ化ファイルの実体が第１ストレージ装置１０ａに格納されていた場合、第１サーバ装置３ａは、第１ストレージ装置１０ａに格納されている当該スタブ化ファイルの実体を更新要求の内容に従って更新するとともに、当該スタブ化ファイルの実体同期要フラグ２３１３をＯＮに設定する（Ｓ３２１３）。

【０１３９】

一方、スタブ化ファイルの実体が第１ストレージ装置１０ａに格納されていない場合、第１サーバ装置３ａは、第２サーバ装置３ｂに当該スタブ化ファイルの実体の取得要求(リコール要求）を送信する（Ｓ３２１４）。

【０１４０】

第１サーバ装置３ａは、上記要求に応じて第２サーバ装置３ｂから送られてきたファイルの実体を受信すると（Ｓ３２１５）、受信した実体の内容を更新要求の内容に従って更新し、更新後の実体を当該スタブ化ファイルの実体として第１ストレージ装置１０ａに格納する。また第１サーバ装置３ａは、当該スタブ化ファイルのスタブ化フラグ２３１１をＯＦＦ、レプリケーションフラグ２３１４をＯＦＦ、メタデータ同期要フラグ２３１２をＯＮに夫々設定する（Ｓ３２１６）。

【０１４１】

＜障害回復時の処理＞

次に第１サーバ装置３ａにおいて何らかの障害が発生して情報処理システム１の機能が停止し、その後、第１サーバ装置３ａが修復されて情報処理システム１の機能を再開させる場合に当該情報処理システム１において行われる処理について説明する。

【０１４２】

図３３は、修復した第１サーバ装置３ａに仮想マシン３１０を復旧させる際に情報処理システム１において行われる処理（以下、仮想マシン復旧処理Ｓ３３００と称する。）を説明する図である。以下、同図とともに仮想マシン復旧処理Ｓ３３００について説明する。

【０１４３】

尚、仮想マシン復旧処理Ｓ３３００を実行する前提として、仮想マシン３１０を復旧させるための仮想マシンイメージ（仮想化制御部３０５に仮想マシン３１０を実現させるために必要な構成情報であり、例えば、ＣＰＵやメモリ等のハードウエア構成、記憶領域の

サイズ、ネットワーク仕様等の情報が含まれる。）が、第２ストレージ装置１０ｂに予め格納されているものとする。

【０１４４】

同図に示すように、まず第１サーバ装置３ａにおいて、ブートローダ等を用いて記録媒体３３１０等に記録されているインストールプログラムを実行し、第１サーバ装置３ａに仮想化制御部３０５をインストールし（Ｓ３３１１）、仮想化制御部３０５の機能を開始させる（Ｓ３３１２）。

【０１４５】

次に、機能を開始した仮想化制御部３０５が、第２サーバ装置３ｂに仮想マシンイメージの提供を要求する（Ｓ３３１３）。

【０１４６】

第２サーバ装置３ｂは、第１サーバ装置３ａから上記要求を受信すると、上記要求に指定されている仮想マシンイメージを第２ストレージ装置１０ｂから取得し（Ｓ３３１４）、取得した仮想マシンイメージを第１サーバ装置３ａに送信する（Ｓ３３１５）。

【０１４７】

尚、第２サーバ装置３ｂは、仮想マシンイメージを、例えば、第１サーバ装置３ａの識別子（以下、サーバＩＤと称する。）と第１サーバ装置３ａにおいて実現される仮想マシン３１０の識別子（以下、仮想マシンＩＤと称する。）とに対応付けて管理しており、上記取得要求を受信すると、その取得要求に指定されているサーバＩＤ及び仮想マシンＩＤで特定される仮想マシンイメージを特定し、特定した仮想マシンイメージを第１ストレージ装置１０ａに送信する。

【０１４８】

第１サーバ装置３ａは、第２サーバ装置３ｂから仮想マシンイメージを受信すると（Ｓ３３１６）、受信した仮想マシンイメージを第１ストレージ装置１０ａに格納し、受信した仮想マシンイメージに基づく仮想マシン３１０の動作を開始させる（Ｓ３３１７）。

【０１４９】

尚、以上に説明した仮想マシン復旧処理Ｓ３３００は、原則として仮想マシンイメージに基づく仮想マシン３１０の再起動が必要になるような大きな障害が発生した場合に行われ、例えば、障害が仮想マシン３１０の再起動が必要になるような障害でない場合は必ずしも仮想マシン３１０を再起動させる必要はない。

【０１５０】

図３４は、図３３に示した仮想マシン復旧処理Ｓ３３００により第１サーバ装置３ａにおいて仮想マシン３１０が動作を開始した後、クライアント装置２からのデータＩ／Ｏ要求を受け付ける前に、情報処理システム１において行われる、ディレクトリイメージを復旧させる処理（以下、ディレクトリイメージ事前回復処理Ｓ３４００と称する。）を説明する図である。以下、同図とともにディレクトリイメージ事前回復処理Ｓ３４００について説明する。

【０１５１】

まず第１サーバ装置３ａが、仮想マシン復旧処理Ｓ３３００によって再起動された仮想マシン３１０のファイルシステム３１２が、障害が発生する前に第１ストレージ装置１０ａに構成していたディレクトリの構成（つまり第２ストレージ装置１０ｂに記憶されているディレクトリの構成であり、ディレクトリの階層構造を示すデータ、ディレクトリのデータ（メタデータ）、及びファイルのデータ（メタデータ及び実体)を含む。以下、ディレクトリイメージと称する。）における、最上位ディレクトリ（以下、ルートディレクトリと称する。）に存在するディレクトリのメタデータ、及びルートディレクトリに存在するファイルのメタデータの取得要求を、第２サーバ装置３ｂに送信する（Ｓ３４１１）。

【０１５２】

尚、本実施形態において、ルートディレクトリに存在するディレクトリのメタデータ、及びルートディレクトリに存在するファイルのメタデータという場合には、ルートディレクトリに存在する（から見える）ディレクトリ及びファイルは含まれるが、ルートディレ

クトリに存在するディレクトリの更に配下に存在するディレクトリやそのディレクトリに存在するファイルは含まれない。

【０１５３】

第２サーバ装置３ｂは、上記取得要求を受信すると、要求されているルートディレクトリに存在するディレクトリのメタデータ、及びルートディレクトリに存在するファイルのメタデータを、第２ストレージ装置１０ｂから取得し（Ｓ３４１２）、取得したメタデータを第１ストレージ装置１０ａに送信する（Ｓ３４１３）。

【０１５４】

尚、第２サーバ装置３ｂは、前述したレプリケーションによる管理方式の運用において、メタデータをサーバＩＤと仮想マシンＩＤとに対応づけて管理し、第２サーバ装置３ｂは、上記取得要求を受信すると、その取得要求に指定されているサーバＩＤ及び仮想マシンＩＤで特定されるメタデータを特定し、特定したメタデータを第２ストレージ装置１０ｂから取得する。

【０１５５】

第１サーバ装置３ａは、第２サーバ装置３ｂからメタデータを受信すると（Ｓ３４１３）、受信したメタデータに基づくディレクトリイメージを、第１ストレージ装置１０ａに復元する（Ｓ３４１４）。またこのとき、第１サーバ装置３ａは、メタデータ同期要フラグ２３１２をＯＮ、実体同期要フラグ２３１３をＯＮに夫々設定する。尚、復元されたファイルはいずれもメタデータのみに基づくものであるため、これらのファイルはいずれもスタブ化された状態になっており、スタブ化フラグ２３１１がＯＮに設定されている。

【０１５６】

そして第１サーバ装置３ａは、第１ストレージ装置１０ａにディレクトリイメージが復元されると、クライアント装置２へのサービスを開始する。

【０１５７】

図３５は、図３４に示したディレクトリイメージ事前回復処理Ｓ３４００の後、クライアント装置２からのデータＩ／Ｏ要求の受け付けを開始した第１サーバ装置３ａが、障害発生前に当該第１サーバ装置３ａが管理していたディレクトリイメージを復元していく処理（以下、オンデマンド復元処理Ｓ３５００と称する。）を説明する図である。以下、同図とともにオンデマンド復元処理Ｓ３５００について説明する。

【０１５８】

第１サーバ装置３ａは、サービスを開始した後、クライアント装置２からあるファイルについてデータＩ／Ｏ要求を受け付けると（Ｓ３５１１）、受け付けたデータＩ／Ｏ要求の対象になっているファイル（以下、アクセス対象ファイルと称する。）のメタデータが、第１ストレージ装置１０ａに存在するか否か（サービス開始後、既に第１ストレージ装置１０ａにメタデータが復元されているか否か）を調べる（Ｓ３５１２）。

【０１５９】

メタデータが第１ストレージ装置１０ａに復元されている場合、第１サーバ装置３ａは、受け付けたデータＩ／Ｏ要求の対象（メタデータ又は実体）、データＩ／Ｏ要求の種類（参照要求か更新要求か）、レプリケーションによる管理方式で管理されているか否か（レプリケーションフラグ２３１４がＯＮか否か）、スタブ化されているか否か（スタブ化フラグがＯＮか否か）に応じて、受け付けたデータＩ／Ｏ要求に対応する処理（前述したレプリケーションファイル更新処理Ｓ２７００、レプリケーションファイル参照処理Ｓ２８００、メタデータアクセス処理Ｓ３０００、スタブ化ファイル実体参照処理Ｓ３１００、スタブ化ファイル実体更新処理Ｓ３２００）を行い、クライアント装置２に応答を返す（Ｓ３５１８）。

【０１６０】

一方、アクセス対象ファイルのメタデータが復元されていない場合には、第１サーバ装置３ａは、ルートディレクトリを起点としてアクセス対象ファイルが存在するディレクトリレベル（ディレクトリ階層）に至るまでのディレクトリイメージを復元するためのデータを第２サーバ装置３ｂ（第２ストレージ装置１０ｂ）から取得し（Ｓ３５１３～Ｓ３５

１５）、取得したデータを用いてルートディレクトリから上記ディレクトリレベルに至るまでのディレクトリイメージを第１ストレージ装置１０ａに復元する（Ｓ３５１６）。

【０１６１】

また第１サーバ装置３ａは、アクセス対象ファイルのスタブ化フラグ２３１１をＯＮに、レプリケーションフラグ２３１４をＯＦＦに、メタデータ同期要フラグ２３１２をＯＮに、夫々設定する（Ｓ３５１７）。

【０１６２】

次に第１サーバ装置３ａは、受け付けたデータＩ／Ｏ要求の対象や種類、管理方式、スタブ化の有無等に応じて、受け付けたデータＩ／Ｏ要求に対応する処理を行い、クライアント装置２に応答を返す（Ｓ３５１８）。

【０１６３】

図３６に、データＩ／Ｏ要求が繰り返し発生することにより、以上に説明したオンデマンド復元処理Ｓ３５００によって段階的に第１ストレージ装置１０ａにディレクトリイメージが復元されていく様子を示している。

【０１６４】

同図中、強調された文字列（アンダーラインが付与されている文字列）で示すディレクトリは、そのディレクトリのメタデータは復元されているがその配下のディレクトリのメタデータは未だ復元されていない。また強調されていない文字で示すディレクトリは、そのディレクトリの配下のディレクトリのメタデータも既に復元されている。また強調された文字で示すファイルは、そのファイルのメタデータは復元されているが、その実体は未だ復元されていない。また強調されていない文字で示すファイルは、そのファイルの実体が既に復元されている。

【０１６５】

図３６の図（０）は、障害が発生する直前に第１サーバ装置３ａ（第１ストレージ装置１０ａ）において管理されていたディレクトリイメージ（最終的に復元されるディレクトリイメージの全体）である。

【０１６６】

図３６の図（Ａ）は、ディレクトリイメージ事前回復処理Ｓ３４００による回復直後（まだ第１サーバ装置３ａがデータＩ／Ｏ要求を受け付けていない状態）におけるディレクトリイメージである。この段階ではルートディレクトリ「／」の直下に存在するディレクトリ「／ｄｉｒ１」及び「／ｄｉｒ２」のメタデータは復元されているが、その更に配下のディレクトリのメタデータについては未だ復元されていない。またルートディレクトリ「／」の直下に存在するファイル「ａ．ｔｘｔ」のメタデータは復元されているが、実体については未だ復元されていない。

【０１６７】

図３６の図（Ｂ）は、図（Ａ）の状態でクライアント装置２からディレクトリ「／ｄｉｒ１」の配下に存在するファイル「ｃ．ｔｘｔ」に対するデータＩ／Ｏ要求を受け付けた後の状態である。クライアント装置２からファイル「ｃ．ｔｘｔ」に対するデータＩ／Ｏ要求を受け付けたため、ディレクトリ「／ｄｉｒ１１」のメタデータと「／ｃ．ｔｘｔ」のメタデータが復元されている。

【０１６８】

図３６の図（Ｃ）は、図（Ｂ）の状態で更にクライアント装置２からディレクトリ「／ｄｉｒ２」の配下に存在するファイル「ｂ．ｔｘｔ」に対するデータＩ／Ｏ要求を受け付けた後の状態である。同図に示すように、クライアント装置２からファイル「ｂ．ｔｘｔ」に対するデータＩ／Ｏ要求を受け付けたため、「／ｂ．ｔｘｔ」のメタデータが復元されている。尚、「／ｄｉｒ２」の配下に存在する「／ｂ．ｔｘｔ」のメタデータが復元されたため、「／ｄｉｒ２」は強調されていない文字で表記している。

【０１６９】

図３６の図（Ｄ）は、図（Ｃ）の状態でクライアント装置２からファイル「ｂ．ｔｘｔ」に対するデータＩ／Ｏ要求（更新要求）を受け付けた後の状態である。クライアント装

置２からファイル「ｂ．ｔｘｔ」に対するデータＩ／Ｏ要求（更新要求）を受け付けたため、ファイル「ｂ．ｔｘｔ」の実体が復元されている。

【０１７０】

以上に説明したように、本実施形態の情報処理システム１においては、第１サーバ装置３ａに障害が発生した後、データＩ／Ｏ要求の受け付けを開始した時点では、ディレクトリイメージ事前回復処理Ｓ３４００によってルートディレクトリに存在するディレクトリのメタデータ、及びルートディレクトリに存在するファイルのメタデータのみが復元される。そしてその後は第１サーバ装置３ａに対してクライアント装置２からまだ復元されていないファイルに対してデータＩ／Ｏ要求が発生する度に、段階的に第１サーバ装置３ａ（第１ストレージ装置１０ａ）にディレクトリイメージが復元されていく。

【０１７１】

このように障害回復後、データＩ／Ｏ要求の受け付けを開始する前にディレクトリイメージの全体を復元してしまうのではなく、ディレクトリイメージを段階的に復元するようにすることで、サービスが再開される前にディレクトリイメージの全体を復元しておくようにしておく場合に比べ、障害発生からサービス再開までに要する時間を短縮することができ、ユーザの業務等への影響を防ぐことができる。

【０１７２】

またディレクトリイメージが完全に復元されるまでの間は、第１ストレージ装置１０ａの資源を節約することができる。さらにディレクトリイメージが完全に復元されるまでの間は記憶容量の消費が抑えられるため、例えば、記憶容量の小さなストレージ装置を障害が発生した第１ストレージ装置１０ａの代替装置として用いるようなこともできる。

【０１７３】

＜復元対象の追加＞

ところで、例えば、第１サーバ装置３ａや第１ストレージ装置１０ａが充分な性能や記憶容量を備えている場合や、ユーザがサービスの迅速に完全復旧されることを望んでいるような場合には、図３５に示したオンデマンド復元処理Ｓ３５００によって障害発生前における第１ストレージ装置１０ａのディレクトリイメージを迅速に復元することが好ましい。

【０１７４】

しかし前述したオンデマンド復元処理Ｓ３５００によるディレクトリイメージの復元速度は、クライアント装置２からのデータＩ／Ｏ要求の発生頻度に依存しているため、データＩ／Ｏ要求の発生頻度が少ない場合はディレクトリイメージが完全に復元されるまでに長時間を要してしまうことになる。

【０１７５】

そこで本実施形態の情報処理システム１は、このようなディレクトリイメージの復旧速度の低下を防ぐべく、オンデマンド復元処理Ｓ３５００において、第２サーバ装置３ｂが、第１サーバ装置３ａから復元のためのディレクトリイメージを要求された際、クライアント装置２から受け付けたデータＩ／Ｏ要求が所定の条件を満たすことを条件として、第１サーバ装置３ａに送信するディレクトリイメージを追加し、ディレクトリイメージの復元を自動的に促進させる仕組みを備えている。

【０１７６】

尚、上記の所定の条件としては、例えば次のようなものが考えられる。

【０１７７】

（条件１）アクセス対象ファイルのデータサイズが、現在から遡って所定時間内に発生したデータＩ／Ｏ要求のアクセス対象ファイルのデータサイズの平均値よりも小さい。

【０１７８】

（条件２）アクセス対象ファイルのデータサイズが予め設定された閾値よりも小さい。

【０１７９】

また上記の仕組みにおいて追加するディレクトリイメージの選出方法としては、例えば次のようなものが考えられる。

【０１８０】
（選出方法１）既に復元されているディレクトリの配下に存在するファイルのメタデータ及び／又は実体を選出する。
【０１８１】
ここで一般に既に復元されているディレクトリの配下に存在するファイルはその後にアクセスされる可能性が高く、選出方法１に従いそのようなディレクトリの配下に存在するファイルのディレクトリイメージを復元することで、クライアント装置２に対する応答性能の向上が期待できる。
【０１８２】
（選出方法２）既に復元されているディレクトリの配下に存在するディレクトリのメタデータを選出する。
【０１８３】
ここで既に復元されているディレクトリの配下に存在するディレクトリはその後にアクセスされる可能性が高いので、選出方法２に従い既に復元されているディレクトリの配下に存在するディレクトリのメタデータを復元しておくことでクライアント装置２に対する応答性能の向上が期待できる。
【０１８４】
（選出方法３）第１サーバ装置３ａに障害が発生する前に第１ストレージ装置１０ａに実体が格納されていたファイルの実体（障害が発生する前にスタブ化フラグがＯＦＦになっていたファイル）を選出する。
【０１８５】
第１サーバ装置３ａに障害が発生する前に第１ストレージ装置１０ａに実体が格納されていたファイルは元々アクセス頻度が高いファイルであった可能性が高い。そこでそのようなファイルの実体を優先して第１ストレージ装置１０ａに復元しておくようにすればクライアント装置２に対する応答性能の向上を期待することができる。
【０１８６】
尚、第１サーバ装置３ａは、障害が発生する前に第１ストレージ装置１０ａに実体が格納されていたファイルであるか否かを、例えば、第２サーバ装置３ｂにそのファイルのスタブ化フラグ２３１１を問い合わせることにより把握する（そのファイルのスタブ化フラグ２３１１がＯＦＦであれば障害が発生する前に第１ストレージ装置１０ａに実体が格納されていたことになる）。
【０１８７】
（選出方法４）アクセス対象ファイルよりも高い重要度が設定されているファイルのメタデータ及び／又は実体を選出する。
【０１８８】
一般に高い重要度が設定されているファイルはクライアント装置２からアクセスされる可能性が高いファイルであることが多い。そこでそのようなファイルのメタデータ及び／又は実体を復元しておくようにすることで、クライアント装置２に対する応答性能の向上を期待することができる。
【０１８９】
尚、第１サーバ装置３ａは、第１サーバ装置３ａから第２サーバ装置３ｂに問い合わせを行うことにより、第１ストレージ装置１０ａにメタデータが未だ復元されていないファイルの重要度（ｉｎｏｄｅ管理テーブル１９１２の重要度２３１６の内容）を取得する。

【０１９０】
（選出方法５）障害発生時点から遡って所定時間内におけるアクセス頻度がアクセス対象ファイルよりも高いファイルを選出する。
【０１９１】
障害発生時点から遡って所定時間内におけるアクセス頻度が高いファイルは、クライアント装置２からアクセスされる可能性が高いファイルであると考えられる。そこでそのよ

うなファイルのメタデータ及び／又は実体を復元しておくようにすることで、クライアント装置２に対する応答性能の向上を期待することができる。

【０１９２】

尚、第１サーバ装置３ａは、第２サーバ装置３ｂにファイルアクセスログ３６８の内容を問い合わせることにより、障害発生時点から遡って所定時間内におけるファイルのアクセス頻度を取得する。

【０１９３】

尚、以上に列挙した方法は選出方法の一例に過ぎず、選出方法は以上に示したものに限定されない。例えば以上に列挙した選出方法の二つ以上を組み合わせて復元するディレクトリイメージを選出するようにしてもよい。例えば単独の選出方法だと選出される復元対象が多すぎてしまう場合には、複数の選出方法を組み合わせることで復元対象を絞り込むことができる。

【０１９４】

図３７は、前述したオンデマンド復元処理Ｓ３５００において、データＩ／Ｏ要求が前述した所定の条件を満たす場合に前述した所定の選出方法に従い復元するディレクトリイメージを追加するようにした処理（以下、オンデマンド復元処理（復元対象追加有）Ｓ３７００と称する。）を説明する図である。以下、同図とともにオンデマンド復元処理（復元対象追加有）Ｓ３７００について説明する。

【０１９５】

第１サーバ装置３ａは、クライアント装置２からデータＩ／Ｏ要求を受け付けると（Ｓ３７１１）、当該データＩ／Ｏ要求のアクセス対象ファイルのメタデータが第１ストレージ装置１０ａに存在するか否か（既に復元されているか否か）を判断する（Ｓ３７１２）。

【０１９６】

アクセス対象ファイルのメタデータが既に復元されている場合には、第１サーバ装置３ａは、受け付けたデータＩ／Ｏ要求の対象や種類、管理方式、スタブ化の有無等に応じて、受け付けたデータＩ／Ｏ要求に対応する処理を行い、クライアント装置２に応答を返す（Ｓ３７１８）。

【０１９７】

一方、アクセス対象ファイルのメタデータが復元されていない場合には、第１サーバ装置３ａは、ルートディレクトリを起点としてアクセス対象ファイルが存在するディレクトリレベル（ディレクトリ階層）に至るまでのディレクトリイメージを復元するためのデータを第２サーバ装置３ｂに要求する（ここまでの処理は図３５に示したオンデマンド復元処理Ｓ３５００と同様である。）。

【０１９８】

第２サーバ装置３ｂは、上記要求を受信すると、データＩ／Ｏ要求が前述した所定の条件を満たすか否かを判断し、所定の条件を満たす場合には、前述した所定の選出方法に従い追加するディレクトリイメージをさらに選出する。そしてサーバ装置３ｂは、上記要求に指定されているディレクトリイメージを復元するためのデータと、上記選出したディレクトリイメージを復元するためのデータとを、第２ストレージ装置１０ｂから取得し、第１サーバ装置３ａに送信する（Ｓ３７１３～Ｓ３７１５）。

【０１９９】

第１サーバ装置３ａは、第２サーバ装置３ｂから上記データを受信すると、受信したデータを用いて、第１ストレージ装置１０ａにディレクトリイメージを復元する（Ｓ３７１６）。

【０２００】

次に第１サーバ装置３ａは、アクセス対象ファイルのスタブ化フラグ２３１１をＯＮに、レプリケーションフラグ２３１４をＯＦＦに、メタデータ同期要フラグ２３１２をＯＮに、夫々設定する（Ｓ３７１７）。

【０２０１】

そして第１サーバ装置３ａは、受け付けたデータＩ／Ｏ要求の対象や種類、管理方式、スタブ化の有無等に応じて、受け付けたデータＩ／Ｏ要求に対応する処理を行い、クライアント装置２に応答を返す（Ｓ３７１８）。

【０２０２】

以上に説明したオンデマンド復元処理（復元対象追加有）Ｓ３７００によれば、データＩ／Ｏ要求が所定の条件を満たす場合は、復元されるディレクトリイメージが自動的に追加される。このため、ディレクトリイメージの復旧速度を自動的に速めることができ、第１ストレージ装置１０ａのディレクトリイメージを迅速に障害発生前の状態に復元することができる。

【０２０３】

また以上に説明したオンデマンド復元処理（復元対象追加有）Ｓ３７００では、復元するディレクトリイメージを追加するか否かの判断やディレクトリイメージを復元するためのデータの取得に関する処理を、もっぱら第２サーバ装置３ｂ側で行うようにしている。そのため、第１サーバ装置３ａ側に特別な仕組みを設ける必要がなく、また第１サーバ装置３ａの代替装置の選出に際し、機種や製造元（ベンダー）を一致させる必要がない等、情報処理システム１の柔軟な運用形態が可能になる。

【０２０４】

＜再スタブ化の回避＞

前述したスタブ化候補選出処理Ｓ２５００（図２５）では、ファイル格納領域の残容量がスタブ化閾値未満になったことを条件としてスタブ化の候補となるファイルが選出され、選出されたファイルは前述したスタブ化処理Ｓ２６００（図２６）において実際にスタブ化（第１ストレージ装置１０ａから実体が削除）されるが、スタブ化候補選出処理Ｓ２５００及びスタブ化処理Ｓ２６００は、図３５に説明したオンデマンド復元処理Ｓ３５００（又は図３７に示したオンデマンド復元処理（復元対象追加有）Ｓ３７００。以下、オンデマンド復元処理Ｓ３５００のみ表記する。）の実行中においても実行されてしまうことがある。

【０２０５】

そして例えばスタブ化閾値が高めに設定されていた場合や、障害が発生した第１ストレージ装置１０ａの代替として用意されたストレージ装置が充分な記憶容量を備えていない場合には、オンデマンド復元処理Ｓ３５００によって第１ストレージ装置１０ａに実体が復元されたファイルが、オンデマンド復元処理Ｓ３５００（又は図３７に示したオンデマンド復元処理（復元対象追加有）Ｓ３７００）によって直ぐに再びスタブ化の候補として選出されスタブ化されてしまう（以下、この現象のことを再スタブ化と称する。）。

【０２０６】

そしてこのような再スタブ化が頻発すると、情報処理システム１の資源が浪費され、情報処理システム１の運用効率が低下してしまうことになる。

【０２０７】

そこで本実施形態の情報処理システム１は、再スタブ化の発生を抑制すべく、再スタブ化の発生を随時監視し、再スタブ化の発生状況に応じてディレクトリイメージの復元を自動的に抑制する仕組みを備えている。

【０２０８】

図３８は、上記仕組みに関して第２サーバ装置３ｂによって行われる処理（以下、再スタブ化回避処理Ｓ３８００と称する。）を説明する図である。以下、同図とともに再スタブ化回避処理Ｓ３８００について説明する。

【０２０９】

第２サーバ装置３ｂは、前述したオンデマンド復元処理Ｓ３５００の実行中、単位時間当たりの再スタブ化の発生頻度が予め設定された閾値（以下、再スタブ化頻度閾値と称する。）以上になっているか否か、もしくは再スタブ化の発生時間間隔が予め設定された閾値（以下、再スタブ化発生時間間隔閾値と称する。）未満になっているか否かを監視する（Ｓ３８１１～Ｓ３８１３）。

【０２１０】
ここで再スタブ化が発生したか否かの判断は、例えば、リストアログ３６５の内容と、メタデータの同期処理Ｓ２９００における第１サーバ装置３ａから第２サーバ装置３ｂへのスタブ化フラグ２３１１の更新通知（スタブ化フラグ２３１１をＯＦＦからＯＮにする通知））とに基づいて行う。
【０２１１】
例えば、第２サーバ装置３ｂは、第１ストレージ装置１０ａにディレクトリイメージが復元されてから予め設定された所定時間内に、そのディレクトリイメージのデータ（メタデータ又は実体）のスタブ化フラグ２３１１がＯＮにされたことをもって再スタブ化が発生したと判断する。
【０２１２】
同図に示すように、第２サーバ装置３ｂは、上記監視において再スタブ化の発生頻度が再スタブ化頻度閾値以上になっていること、もしくは再スタブ化の発生時間間隔が再スタブ化発生時間間隔閾値未満になっていることを検知すると、第１サーバ装置３ａに送信するディレクトリイメージ(図３７に示したオンデマンド復元処理（復元対象追加有）Ｓ３７００において追加されるディレクトリイメージを含む）の量を抑制（減少）させる。尚、この抑制には、第１サーバ装置３ａへのディレクトリイメージの送信を中止する場合も含まれる（Ｓ３８１４）。
【０２１３】
ここで上記抑制の具体的な方法としては、例えば、次のようなものが考えられる。
【０２１４】
（抑制方法１）データＩ／Ｏ要求がメタデータのみを対象としている場合に当該ファイルの実体を復元しないようにする。
【０２１５】
これによれば実体を復元するのに要する負荷を軽減することができる。またデータＩ／Ｏ要求がメタデータのみを対象としている場合には、当該ファイルの実体まで復元する必要がないため、実体を復元しなくてもデータＩ／Ｏ要求の処理に影響を与えることもない。
【０２１６】
（抑制方法２）前述した（選出方法１）～（選出方法５）の一つ以上を用いたディレクトリイメージの選出を行っている場合に、更に別の選出方法を重複して適用するようにする。
【０２１７】
選出方法を重複して適用するようにすることで、段階的に再スタブ化の発生を抑制することができ、再スタブ化の発生状況に併せて、第１サーバ装置３ａに送信するディレクトリイメージの量を適切に抑制することができる。
【０２１８】
（抑制方法３）前述した（選出方法４）において判断に用いている重要度の閾値を更に高く設定する。
【０２１９】
重要度の閾値を更に高く設定することで、再スタブ化の抑制を容易に実施することができる。また重要度の閾値を段階的に高く設定していくようにすれば、再スタブ化の発生状況に併せて、第１サーバ装置３ａに送信するディレクトリイメージの量を適切に抑制することができる。
【０２２０】
（抑制方法４）前述した（選出方法５）においてアクセス頻度の判断に用いているアクセス頻度の閾値を更に高く設定する。
【０２２１】
アクセス頻度の閾値を更に高く設定することで、再スタブ化の抑制を容易に実施することができる。またアクセス頻度の閾値を段階的に高く設定していくようにすれば、再スタ

ブ化の発生状況に併せて、第１サーバ装置３ａに送信するディレクトリイメージの量を適切に抑制することができる。

【０２２２】

尚、第２サーバ装置３ｂは上記監視を継続して行い、再スタブ化の発生頻度が再スタブ化頻度閾値未満になるとともに、再スタブ化の発生時間間隔が再スタブ化発生時間間隔閾値以上になると、自動的に上記抑制を解除する。ここで解除には、抑制を一度に全て解除してしまう場合のほか、少しずつディレクトリイメージを追加していく等、抑制を段階的に解除していく場合も含む（Ｓ３８１４）。

【０２２３】

以上に説明したように、再スタブ化回避処理Ｓ３８００によれば、再スタブ化が頻繁に発生すると第２サーバ装置３ｂから第１サーバ装置３ａに送信されるディレクトリイメージの量が自動的に抑制されるので、再スタブ化の発生を抑制することができる。そのため、再スタブ化による情報処理システム１の資源の浪費を防ぐことができ、再スタブ化に起因する情報処理システム１を運用効率の低下を防ぐことができる。

【０２２４】

また以上に説明した再スタブ化回避処理Ｓ３８００は、第２サーバ装置３ｂが主体となって行うため、第１サーバ装置３ａ側に特別な仕組みを設ける必要がない。そのため、再スタブ化を抑制するための仕組みを情報処理システム１に容易に実現することができる。また特別な性能や仕様が要求されないことで第１ストレージ装置１０ａの選択肢が拡がり、ベンダーや機種等に依存することなく、ハードウエア及びソフトウエアを自由に選択することができる。

【０２２５】

＜処理詳細＞

次に情報処理システム１において行われる処理の詳細について説明する。

【０２２６】

図３９は、図２４に示したレプリケーション開始処理Ｓ２４００の詳細を説明するフローチャートである。以下、同図とともに説明する。

【０２２７】

第１サーバ装置３ａは、クライアント装置２等からレプリケーション開始要求を受け付けたか否かをリアルタイムに監視している（Ｓ３９１１）。第１サーバ装置３ａは、クライアント装置２等からレプリケーション開始要求を受け付けると（Ｓ３９１１：ＹＥＳ）（図２４のＳ２４１１）、受け付けたレプリケーション開始要求に指定されているファイルのデータ（メタデータ及び実体）の格納先（ＲＡＩＤグループの識別子、ブロックアドレス等）を第２サーバ装置３ｂに問い合わせる（Ｓ３９１２）。

【０２２８】

第２サーバ装置３ｂは、上記の問い合わせがあると（Ｓ３９２１）、第２ストレージ装置１０ｂの空き領域を探索してファイルのデータの格納先を決定し、決定した格納先を第１サーバ装置３ａに通知する（Ｓ３９２２）。

【０２２９】

第１サーバ装置３ａは、上記通知を受信すると（Ｓ３９１３）、受け付けたレプリケーション開始要求に指定されているファイルのデータ（メタデータ及び実体）を、第１ストレージ装置１０ａから読み出し（Ｓ３９１４）（図２４のＳ２４１２）、読み出したファイルのデータをＳ３９２２で通知された格納先とともに第２サーバ装置３ｂに送信する（Ｓ３９１５）（図２４のＳ２４１３）。

【０２３０】

また第１サーバ装置３ａは、当該ファイルのメタデータ（第１ストレージ装置１０ａに格納されている当該ファイルのメタデータ）のレプリケーションフラグ２３１４をＯＮ、メタデータ同期要フラグ２３１２をＯＮに夫々設定する（Ｓ３９１６）（図２４のＳ２４１４）。

【０２３１】

尚、メタデータ同期要フラグ２３１２をＯＮに設定しておくことで、前述した同期処理Ｓ２９００により第１ストレージ装置１０ａに格納されているファイルのメタデータと、その複製として第２ストレージ装置１０ｂに格納されているファイルのメタデータの一致性が、同期又は非同期に確保（保証）される。

【０２３２】

一方、第２サーバ装置３ｂは、第１サーバ装置３ａからファイルのデータを受信すると（Ｓ３９２３）、受信したファイルのデータを、当該ファイルとともに受信した格納先で特定される第２ストレージ装置１０ｂの位置に格納する（Ｓ３９２４）。

【０２３３】

図４０は、図２５に示したスタブ化候補選出処理Ｓ２５００の詳細を説明するフローチャートである。以下、同図とともに説明する。

【０２３４】

第１サーバ装置３ａは、ファイル格納領域の残容量がスタブ化閾値未満になっているか否かを随時監視し（Ｓ４０１１、Ｓ４０１２）、ファイル格納領域の残容量がスタブ化閾値未満になっていることを検知すると、前述した所定の選出基準に従い第１ストレージ装置１０ａに格納されているレプリケーションファイルの中からスタブ化の候補を選出する（Ｓ４０１２）（図２５のＳ２５１１）。

【０２３５】

そして第１サーバ装置３ａは、スタブ化の候補を選出すると（Ｓ４０１３）、選出したレプリケーションファイルのスタブ化フラグ２３１１をＯＮ、レプリケーションフラグ２３１４をＯＦＦ、メタデータ同期要フラグ２３１２をＯＮに夫々設定する（Ｓ４０１４）（図２５のＳ２５１２）。

【０２３６】

図４１は、図２６に示したスタブ化処理Ｓ２６００の詳細を説明するフローチャートである。以下、同図とともに説明する。

【０２３７】

第１サーバ装置３ａは、第１ストレージ装置１０ａのファイル格納領域に格納されているファイルの中から、スタブ化候補として選出されているファイル（スタブ化フラグ２３１１がＯＮに設定されているファイル）を随時抽出する（Ｓ４１１１，Ｓ４１１２）。

【０２３８】

そして第１サーバ装置３ａは、抽出したファイルの実体を第１ストレージ装置１０ａから削除するとともに（Ｓ４１１３）、抽出したファイルのメタデータからそのファイルの第１ストレージ装置１０ａの格納先を示す情報に無効な値を設定し（例えば、メタデータの当該ファイルの格納先を設定する欄（例えばブロックアドレス２２１８）にＮＵＬＬ値やゼロを設定する。）（Ｓ４１１４）、メタデータ同期要フラグ２３１２をＯＮに設定する（Ｓ４１１５）（図２６のＳ２６１１）。

【０２３９】

図４２は、図２７に示したレプリケーションファイル更新処理Ｓ２７００の詳細を説明するフローチャートである。以下、同図とともに説明する。

【０２４０】

第１サーバ装置３ａは、クライアント装置２からレプリケーションファイルに対する更新要求を受け付けたか否かをリアルタイムに監視している（Ｓ４２１１）。第１サーバ装置３ａは、更新要求を受け付けると（Ｓ４２１１：ＹＥＳ）（図２７のＳ２７１１）、第１ストレージ装置１０ａに格納されている、当該更新要求の対象になっているレプリケーションファイルのデータ（メタデータ又は実体）を、受け付けた更新要求に従って更新する（Ｓ４２１２）（図２７のＳ２７１２）。

【０２４１】

また第１サーバ装置３ａは、メタデータを更新した場合はそのレプリケーションファイルのメタデータ同期要フラグ２３１２をＯＮに設定し（Ｓ４２１３）、レプリケーションファイルの実体を更新した場合はそのレプリケーションファイルの実体同期要フラグ２３

１３をＯＮに設定する（Ｓ４２１４）（図２７のＳ２７１３）。
【０２４２】
図４３は、図２８に示したレプリケーションファイル参照処理Ｓ２８００の詳細を説明するフローチャートである。以下、同図とともに説明する。
【０２４３】
第１サーバ装置３ａは、クライアント装置２からレプリケーションファイルに対する参照要求を受け付けたか否かをリアルタイムに監視している（Ｓ４３１１）。第１サーバ装置３ａは、参照要求を受け付けると（Ｓ４３１１：ＹＥＳ）（図２８のＳ２８１１）、そのレプリケーションファイルのデータ（メタデータ又は実体）を第１ストレージ装置１０ａから読み出し（Ｓ４３１２）（図２８のＳ２８１２）、読み出したデータに基づきクライアント装置２に対して応答する情報を生成し、生成した応答情報をクライアント装置２に送信する（Ｓ４３１３）（図２８のＳ２８１３）。
【０２４４】
図４４は、図２９に示した同期処理Ｓ２９００の詳細を説明するフローチャートである。以下、同図とともに説明する。
【０２４５】
第１サーバ装置３ａは、クライアント装置２からレプリケーションファイルの同期要求を受け付けたか否かをリアルタイムに監視している（Ｓ４４１１）。第１サーバ装置３ａは、同期要求を受け付けると（Ｓ４４１１：ＹＥＳ）（図２９のＳ２９１１）、第１ストレージ装置１０ａのファイル格納領域に格納されているレプリケーションファイルのうち、メタデータ同期要フラグ２３１２又は実体同期要フラグ２３１３の少なくともいずれかがＯＮに設定されているファイルを取得する（Ｓ４４１２）（図２９のＳ２９１２）。
【０２４６】
そして第１サーバ装置３ａは、取得したファイルのメタデータ又は実体を第２サーバ装置３ｂに送信するとともに（Ｓ４４１３）、当該レプリケーションファイルのメタデータ同期要フラグ２３１２又は実体同期要フラグ２３１３をＯＦＦに設定する（Ｓ４４１４）（図２９のＳ２９１３）。
【０２４７】
一方、第２サーバ装置３ｂは、メタデータ又は実体を受信すると（Ｓ４４２１）（図２９のＳ２９１３）、受信したメタデータ又は実体に対応する、第２ストレージ装置１０ｂに格納されているファイルのメタデータ又は実体を、受信したメタデータ又は実体（もしくは更新差分）に基づき更新する（Ｓ４４２２）（図２９のＳ２９１４）。
【０２４８】
図４５は、図３０に示したメタデータアクセス処理Ｓ３０００の詳細を説明するフローチャートである。以下、同図とともに説明する。
【０２４９】
第１サーバ装置３ａは、クライアント装置２からスタブ化されているファイルのメタデータに対するアクセス要求（参照要求又は更新要求）を受け付けたか否かをリアルタイムに監視している（Ｓ４５１１）。
【０２５０】
第１サーバ装置３ａは、スタブ化されているファイルのメタデータに対するアクセス要求を受け付けると（Ｓ４５１１：ＹＥＳ）（図３０のＳ３０１１）、受け付けたアクセス要求の対象になっている第１ストレージ装置１０ａのメタデータを取得し（Ｓ４５１２）、受け付けたアクセス要求に従い（Ｓ４５１３）、メタデータの参照（読み出したメタデータに基づく応答情報のクライアント装置２への送信）（Ｓ１５１４）、又はメタデータの更新を行う（Ｓ４５１５）（図３０のＳ３０１２）。またメタデータの内容を更新した場合は（Ｓ４５１５）、そのファイルのメタデータ同期要フラグ２３１２にＯＮを設定する（図３０のＳ３０１３）。
【０２５１】
図４６は、図３１に示したスタブ化ファイル実体参照処理Ｓ３１００の詳細を説明する

フローチャートである。以下、同図とともに説明する。

【０２５２】

第１サーバ装置３ａは、クライアント装置２からスタブ化ファイルの実体に対する参照要求を受け付けると（Ｓ４６１１：ＹＥＳ）（図３１のＳ３１１１）、当該スタブ化ファイルの実体が第１ストレージ装置１０ａに格納されているか否かを判断する（Ｓ４６１２）（図３１のＳ３１１２）。

【０２５３】

スタブ化ファイルの実体が第１ストレージ装置１０ａに格納されている場合には（Ｓ４６１２：ＹＥＳ）、第１サーバ装置３ａは、第１ストレージ装置１０ａから当該スタブ化ファイルの実体を読み出し、読み出した実体に基づきクライアント装置２に応答する情報を生成し、生成した応答情報をクライアント装置２に送信する（Ｓ４６１３）（図３１のＳ３１１３）。

【０２５４】

一方、スタブ化ファイルの実体が第１ストレージ装置１０ａに格納されていない場合には（Ｓ４６１２：ＮＯ）、第１サーバ装置３ａは、第２サーバ装置３ｂに対してスタブ化ファイルの実体を要求する（リコール要求）（Ｓ４６１４）（図３１のＳ３１１４）。

【０２５５】

第１サーバ装置３ａは、上記取得要求に応じて第２サーバ装置３ｂから送られてくるスタブ化ファイルの実体を受信すると（Ｓ４６２１、Ｓ４６２２、Ｓ４６１５）（図３１のＳ３１１５）、受信した実体に基づき応答情報を生成し、生成した応答情報をクライアント装置２に送信する（Ｓ４６１６）（図３１のＳ３１１６）。

【０２５６】

また第１サーバ装置３ａは、上記第２サーバ装置３ｂから受信した実体を第１ストレージ装置１０ａに格納し、当該スタブ化ファイルのメタデータの当該ファイルの実体の格納先を示す情報（例えばブロックアドレス２２１８）に、当該ファイルの第１ストレージ装置１０ａにおける格納先を示す内容を設定する（Ｓ４６１７）。

【０２５７】

また第１サーバ装置３ａは、当該ファイルのスタブ化フラグ２３１１をＯＦＦに、レプリケーションフラグ２３１４をＯＮに、メタデータ同期要フラグ２３１２をＯＮに、夫々設定する（Ｓ４６１８）（図３１のＳ３１１７）。

【０２５８】

図４７は、図３２に示したスタブ化ファイル実体更新処理Ｓ３２００の詳細を説明するフローチャートである。以下、同図とともに説明する。

【０２５９】

第１サーバ装置３ａは、クライアント装置２からスタブ化ファイルの実体に対する更新要求を受け付けると（Ｓ４７１１：ＹＥＳ）（図３２のＳ３２１１）、当該スタブ化ファイルの実体が第１ストレージ装置１０ａに格納されているか否かを判断する（Ｓ４７１２）（図３２のＳ３２１２）。

【０２６０】

スタブ化ファイルの実体が第１ストレージ装置１０ａに格納されている場合（Ｓ４７１２：ＹＥＳ）、第１サーバ装置３ａは、第１ストレージ装置１０ａに格納されている当該スタブ化ファイルの実体を更新要求の内容に従って更新するとともに（Ｓ４７１３）、当該スタブ化ファイルの実体同期要フラグ２３１３をＯＮに設定する（Ｓ４７１４）（図３２のＳ３２１３）。

【０２６１】

一方、上記判断の結果、スタブ化ファイルの実体が第１ストレージ装置１０ａに格納されていない場合には（Ｓ４７１２：ＮＯ）、第１サーバ装置３ａは、第２サーバ装置３ｂに当該スタブ化ファイルの実体の取得要求（リコール要求）を送信する（Ｓ４７１５）（図３２のＳ３２１４）。

【０２６２】

第１サーバ装置３ａは、上記要求に応じて第２サーバ装置３ｂから送られてくるファイルの実体を受信すると（Ｓ４７２１、Ｓ４７２２、Ｓ４７１６）（Ｓ３２１５）、受信した実体の内容を更新要求の内容に従って更新し（Ｓ４７１７）、更新後の実体を当該スタブ化ファイルの実体として第１ストレージ装置１０ａに格納する（Ｓ４７１８）（図３２のＳ３２１６）。

【０２６３】

また第１サーバ装置３ａは、当該スタブ化ファイルのスタブ化フラグ２３１１をＯＦＦ、レプリケーションフラグ２３１４をＯＦＦ、メタデータ同期要フラグ２３１２をＯＮに夫々設定する（Ｓ４７１９）。

【０２６４】

図４８は、図３３に示した仮想マシン復旧処理Ｓ３３００及び図３４に示したディレクトリイメージ事前回復処理Ｓ３４００の詳細を説明するフローチャートである。以下、同図とともに説明する。

【０２６５】

まず第１サーバ装置３ａにおいて、ブートローダ等を用いて記録媒体に記録されているインストールプログラムを実行し、第１サーバ装置３ａに仮想化制御部３０５をインストールし、仮想化制御部３０５の機能を開始させる（Ｓ４８１１）（図３３のＳ３３１１、Ｓ３３１２）。

【０２６６】

次に機能を開始した仮想化制御部３０５が、第２サーバ装置３ｂに仮想マシンイメージの取得要求を送信する（Ｓ４８１２）（図３３のＳ３３１３）。

【０２６７】

第２サーバ装置３ｂは、第１サーバ装置３ａから上記取得要求を受信すると（Ｓ４８２１）、当該取得要求に指定されている仮想マシンイメージを第２ストレージ装置１０ｂから取得し、取得した仮想マシンイメージを第１サーバ装置３ａに送信する（Ｓ４８２２）（図３３のＳ３３１４、Ｓ３３１５）。

【０２６８】

第１サーバ装置３ａは、第２サーバ装置３ｂから仮想マシンイメージを受信すると（Ｓ４８１３）（図３３のＳ３３１６）、受信した仮想マシンイメージを第１ストレージ装置１０ａに格納し（Ｓ４８１４）、受信した仮想マシンイメージに基づき仮想マシン３１０の動作を開始する（Ｓ４８１５）（図３３のＳ３３１７）。

【０２６９】

次に第１サーバ装置３ａは、第２サーバ装置３ｂに対し、仮想マシン復旧処理Ｓ３３００によって再起動された仮想マシン３１０のファイルシステム３１２が障害発生前に構成していたディレクトリイメージのルートディレクトリに存在するディレクトリのメタデータ、及びルートディレクトリに存在するファイルのメタデータの取得要求を、第２サーバ装置３ｂに送信する（Ｓ４８１６）（図３４のＳ３４１１）。

【０２７０】

第２サーバ装置３ｂは、上記取得要求を受信すると（Ｓ４８２３）、要求されているルートディレクトリに存在するディレクトリのメタデータ、及びルートディレクトリに存在するファイルのメタデータを第２ストレージ装置１０ｂから取得し、取得したメタデータを第１ストレージ装置１０ａに送信する（Ｓ４８２４）（図３４のＳ３４１２、Ｓ３４１３）。

【０２７１】

次に第１サーバ装置３ａは、第２サーバ装置３ｂからメタデータを受信すると（Ｓ４８１７）（図３４のＳ３４１３）、受信したメタデータに従ったディレクトリイメージを第１ストレージ装置１０ａに構成（復元）する（Ｓ４８１８）（図３４のＳ３４１４）。またこのとき、第１サーバ装置３ａは、メタデータ同期要フラグ２３１２をＯＮに、実体同期要フラグ２３１３をＯＮに夫々設定する（Ｓ４８１９）。

【０２７２】

そして第１サーバ装置３ａは、第１ストレージ装置１０ａに上記ディレクトリイメージが構成されると、第１サーバ装置３ａはクライアント装置２へのサービスを開始する（Ｓ４８２０）（図３４のＳ３４１５）。

【０２７３】

図４９は、図３５に示したオンデマンド復元処理Ｓ３５００の詳細を説明するフローチャートである。以下、同図とともに説明する。

【０２７４】

第１サーバ装置３ａは、クライアント装置２から、あるファイルについてデータＩ／Ｏ要求を受け付けると（Ｓ４９１１：ＹＥＳ）（図３５のＳ３５１１）、受け付けたデータＩ／Ｏ要求の対象になっているファイル（アクセス対象ファイル）のメタデータが、第１ストレージ装置１０ａに存在するか否かを調べる（Ｓ４９１２）（図３５のＳ３５１２）。

【０２７５】

そしてメタデータが第１ストレージ装置１０ａに復元されている場合（Ｓ４９１２：ＹＥＳ）、第１サーバ装置３ａは、受け付けたデータＩ／Ｏ要求の対象や種類、管理方式、スタブ化の有無等に応じて、受け付けたデータＩ／Ｏ要求に対応する処理を行い、クライアント装置２に応答を返す（Ｓ４９１３）（図３５のＳ３５１８）。

【０２７６】

一方、アクセス対象ファイルのメタデータが第１ストレージ装置１０ａに復元されていない場合には（Ｓ４９１２：ＮＯ）、第１サーバ装置３ａは、ルートディレクトリを起点としてアクセス対象ファイルが存在するディレクトリレベルに至るまでのディレクトリイメージを復元するためのデータを第２サーバ装置３ｂに要求する（Ｓ４９１４）。

【０２７７】

第２サーバ装置３ｂは、要求されたデータを第２ストレージ装置１０ｂから取得し、取得したデータを第１サーバ装置３ａに送信する（Ｓ４９２１、Ｓ４９２２、Ｓ４９１５）。

【０２７８】

第１サーバ装置３ａは、第２サーバ装置３ｂから送られてくるデータを受信すると（Ｓ４９１５）、そのデータを用いてディレクトリイメージを第１ストレージ装置１０ａに復元する（Ｓ４９１６）（図３５のＳ３５１３～Ｓ３５１６）。

【０２７９】

また第１サーバ装置３ａは、アクセス対象ファイルのスタブ化フラグ２３１１をＯＮに、レプリケーションフラグ２３１４をＯＦＦに、メタデータ同期要フラグ２３１２をＯＮに、夫々設定する（Ｓ４９１７）（図３５のＳ３５１７）。

【０２８０】

次に第１サーバ装置３ａは、受け付けたデータＩ／Ｏ要求の対象や種類、管理方式、スタブ化の有無等に応じて、受け付けたデータＩ／Ｏ要求に対応する処理を行い、クライアント装置２に応答を返す（Ｓ４９１８）（図３５のＳ３５１８）。

【０２８１】

図５０及び図５１は、図３７に示したオンデマンド復元処理（復元対象追加有）Ｓ３７００の詳細を説明するフローチャートである。以下、同図とともに説明する。

【０２８２】

第１サーバ装置３ａは、クライアント装置２から、あるファイルについてデータＩ／Ｏ要求を受け付けると（Ｓ５０１１：ＹＥＳ）（図３７のＳ３７１１）、受け付けたデータＩ／Ｏ要求の対象になっているアクセス対象ファイルのメタデータが、第１ストレージ装置１０ａに存在するか否かを調べる（Ｓ５０１２）（図３７のＳ３７１２）。

【０２８３】

メタデータが第１ストレージ装置１０ａに復元されている場合（Ｓ５０１２：ＹＥＳ）、第１サーバ装置３ａは、受け付けたデータＩ／Ｏ要求の対象や種類、管理方式、スタブ化の有無等に応じて、受け付けたデータＩ／Ｏ要求に対応する処理を行い、クライアント

装置２に応答を返す（Ｓ５０１３）（図３７のＳ３７１８）。
【０２８４】
一方、アクセス対象ファイルのメタデータが第１ストレージ装置１０ａに復元されていない場合には（Ｓ５０１２：ＮＯ）、第１サーバ装置３ａは、ルートディレクトリを起点としてアクセス対象ファイルが存在するディレクトリレベルに至るまでのディレクトリイメージを復元するためのデータを、第２サーバ装置３ｂに要求する（Ｓ５０１４）。
【０２８５】
第２サーバ装置３ｂは、上記要求を受信すると、データＩ／Ｏ要求が前述した所定の条件を満たすか否かを判断する（Ｓ５０２２）。
【０２８６】
所定の条件を満たさない場合には（Ｓ５０２２：ＮＯ）、Ｓ５０２４に進む。一方、所定の条件を満たす場合には（Ｓ５０２２：ＹＥＳ）、第２サーバ装置３ｂは、前述した所定の選出方法に従い追加するディレクトリイメージを選出する（Ｓ５０２３）。
【０２８７】
Ｓ５０２４では、第２サーバ装置３ｂは、Ｓ５０２１で受信した要求に指定されているディレクトリイメージを復元するためのデータと、Ｓ５０２３で選出したディレクトリイメージを復元するためのデータとを、第２ストレージ装置１０ｂから取得し、取得したデータを第１サーバ装置３ａに送信する（図３７のＳ３７１３～Ｓ３７１５）。
【０２８８】
第１サーバ装置３ａは、上記データを受信すると（Ｓ５０１５）、受信したデータを用いて、第１ストレージ装置１０ａにディレクトリイメージを復元する（Ｓ５０１６）（図３７のＳ３７１６）。
【０２８９】
次に第１サーバ装置３ａは、アクセス対象ファイルのスタブ化フラグ２３１１をＯＮに、レプリケーションフラグ２３１４をＯＦＦに、メタデータ同期要フラグ２３１２をＯＮに、夫々設定する（Ｓ５０１７）（図３７のＳ３７１７）。
【０２９０】
そして第１サーバ装置３ａは、受け付けたデータＩ／Ｏ要求の対象や種類、管理方式、スタブ化の有無等に応じて、受け付けたデータＩ／Ｏ要求に対応する処理を行い、クライアント装置２に応答を返す（Ｓ５０１８）（図３７のＳ３７１８）。
【０２９１】
図５２は、図３８に示した再スタブ化回避処理Ｓ３８００の詳細を説明するフローチャートである。以下、同図とともに説明する。
【０２９２】
第２サーバ装置３ｂは、オンデマンド復元処理Ｓ３５００（又は図３７に示したオンデマンド復元処理（復元対象追加有）Ｓ３７００）の実行中、単位時間当たりの再スタブ化の発生頻度が予め設定された閾値（以下、再スタブ化頻度閾値と称する。）以上になっているか否か、もしくは再スタブ化の発生時間間隔が予め設定された閾値（以下、再スタブ化発生時間間隔閾値と称する。）未満になっているか否かを監視する（Ｓ５２１１、Ｓ５２１２）（図３８のＳ３８１１～Ｓ３８１３）。
【０２９３】
第２サーバ装置３ｂは、上記監視において再スタブ化の発生頻度が再スタブ化頻度閾値以上になっていることを検知すると（Ｓ５２１１：ＹＥＳ）、抑制フラグ管理テーブル３６６に管理される抑制フラグ３６６１をＯＮに設定する（Ｓ５２１３）。
【０２９４】
また第２サーバ装置３ｂは、上記監視において、再スタブ化の発生時間間隔が再スタブ化発生時間間隔閾値未満になっていることを検知すると（Ｓ５２１２：ＹＥＳ）、抑制フラグ３６６１をＯＮに設定する（Ｓ５２１３）。
【０２９５】
また第２サーバ装置３ｂは、上記監視において、再スタブ化の発生頻度が再スタブ化頻

度閾値未満になっており（Ｓ５２１１：ＮＯ）、かつ、再スタブ化の発生時間間隔が再スタブ化発生時間間隔閾値以上になっている場合は（Ｓ５２１２：ＮＯ）、抑制フラグ３６６１をＯＦＦに設定する（Ｓ５２１４）（図３８のＳ３８１４）。

【０２９６】

Ｓ５２１５では、第２サーバ装置３ｂは、抑制フラグ３６６１がＯＮかＯＦＦかを判断する。抑制フラグ３６６１がＯＮであれば（Ｓ５２１５：ＯＮ）、第２サーバ装置３ｂから第１サーバ装置３ａに送信するディレクトリイメージの量を抑制する処理を開始する（Ｓ５２１６）。尚、既に抑制を開始している場合には抑制を継続する。

【０２９７】

一方、抑制フラグ３６６１がＯＦＦであれば（Ｓ５２１５：ＯＦＦ）、第２サーバ装置３ｂは、第１サーバ装置３ａに送信するディレクトリイメージの量を抑制する処理を終了する（Ｓ５２１７）。尚、既に抑制を終了している場合は抑制なしの状態を維持する。

【０２９８】

以上詳細に説明したように、本実施形態の情報処理システム１にあっては、第１サーバ装置３ａの障害復旧に際し、第２サーバ装置３ｂが、第１サーバ装置３ａがデータＩ／Ｏ要求の受け付けを開始するのに先立ち、第２ストレージ装置１０ｂに記憶しているファイルのデータのうち最上位のディレクトリから所定の下位階層までのディレクトリイメージを第１サーバ装置３ａに送信し、第１サーバ装置３ａが、第２サーバ装置３ｂから送られてくるディレクトリイメージを第１ストレージ装置１０ａに復元した後、データＩ／Ｏ要求の受け付けを再開する。

【０２９９】

このように、本実施形態の情報処理システム１にあっては、第１サーバ装置３ａの障害復旧に際し、障害発生前に第１ストレージ装置１０ａに存在していたディレクトリイメージの全てを復元するのではなく、最上位のディレクトリから所定の下位階層までのディレクトリイメージのみを復元するので、障害発生前に第１ストレージ装置１０ａに存在していた全てのディレクトリイメージを復元する場合に比べて復元に要する時間を短縮することができ、早期にサービスを再開することができる。また全てのディレクトリイメージを復元する場合に比べて情報処理システム１に与える負荷が少なくて済む。

【０３００】

また第２サーバ装置３ｂは、第１サーバ装置３ａから第１ストレージ装置１０ａに復元されていないディレクトリイメージを要求された場合に、要求の対象になっているディレクトリイメージを第２ストレージ装置１０ｂから読み出して送信するとともに、所定の選出方法に従って選出したディレクトリイメージとは異なる追加のディレクトリイメージを第２ストレージ装置１０ｂから読み出して送信し、第１サーバ装置３ａは、第２サーバ装置３ｂから送られてくるディレクトリイメージに基づきデータＩ／Ｏ要求を処理するとともに、ディレクトリイメージ及び第２サーバ装置３ｂから送られてくる追加のディレクトリイメージを第１ストレージ装置１０ａに復元する。

【０３０１】

このように、本実施形態の情報処理システム１にあっては、第２サーバ装置３ｂが、第１サーバ装置３ａから第１ストレージ装置１０ａに復元されていないディレクトリイメージを要求された場合に、要求の対象になっているディレクトリイメージに加えて、所定の選出方法に従って選出したディレクトリイメージとは異なる追加のディレクトリイメージを第２ストレージ装置１０ｂから読み出して送信し、第１サーバ装置３ａは、ディレクトリイメージ及び追加のディレクトリイメージの双方のディレクトリイメージを第１ストレージ装置１０ａに復元するので、ディレクトリイメージの復旧速度を自動的に速めることができる。

【０３０２】

また第２サーバ装置３ｂは、再スタブ化の発生頻度が予め設定された閾値以上になっているか、もしくは、再スタブ化の発生時間間隔が予め設定された閾値未満になっている場合に、第１サーバ装置３ａへのディレクトリイメージ又は追加のディレクトリイメージの

送信を自動的に抑制するので、再スタブ化の発生を抑制することができ、再スタブ化によって情報処理システム１の資源が浪費されてしまうのを防ぐことができる。

【０３０３】

以上、本実施形態について説明したが、上記実施形態は本発明の理解を容易にするためのものであり、本発明を限定して解釈するためのものではない。本発明は、その趣旨を逸脱することなく、変更、改良され得ると共に、本発明にはその等価物も含まれる。

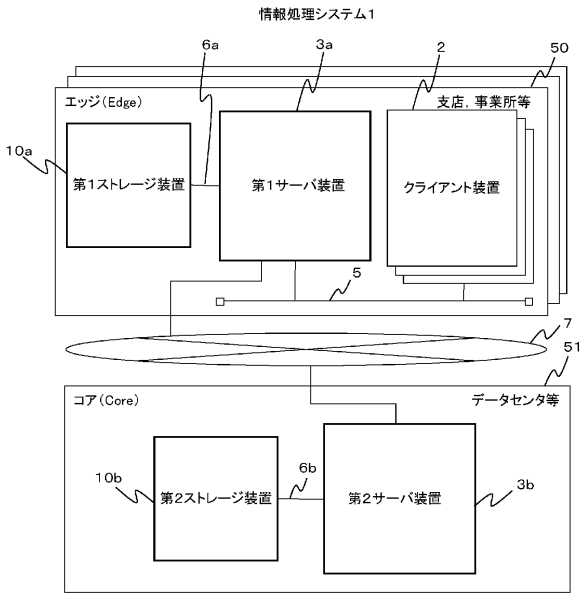
【０３０４】

例えば、以上の説明では、ファイル共有処理部３１１、ファイルシステム３１２、データ操作要求受付部３１３、データ複製／移動処理部３１４、ファイルアクセスログ取得部３１７、及びカーネル／ドライバ３１８の各機能が仮想マシン３１０において実現されているものとして説明したが、これらの機能は必ずしも仮想マシン３１０において実現されるものでなくてもよい。

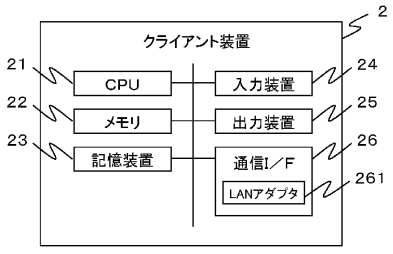
【０３０５】

また前述したディレクトリイメージ事前復旧処理Ｓ３４００では、ファイルシステム３１２が障害発生前に構成していたディレクトリイメージのルートディレクトリに存在するディレクトリのメタデータ、及びルートディレクトリに存在するファイルのメタデータを復旧させるようにしているが、第１サーバ装置３ａの能力に余裕がある場合には、より下位階層までのディレクトリイメージを復旧させるようにしてもよい。

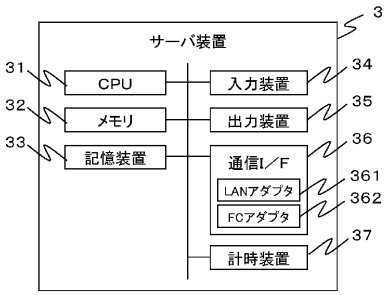
【図１】



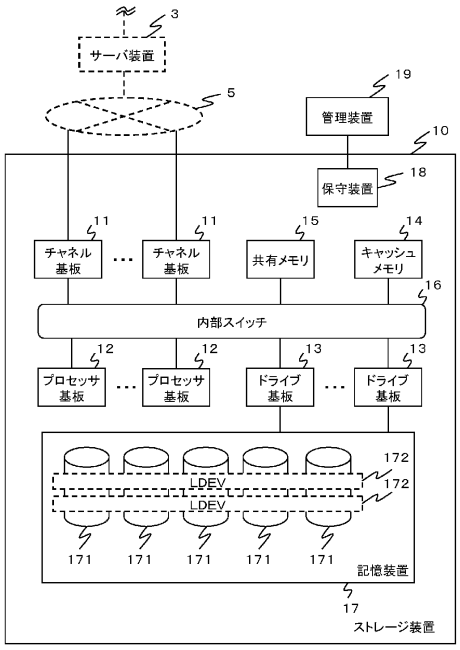
【図２】



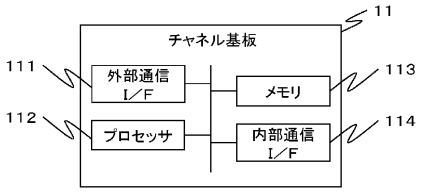
【図３】



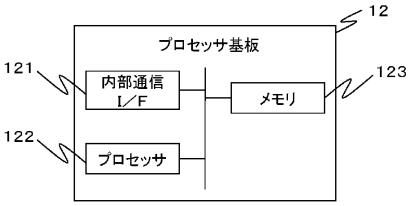
【図４】



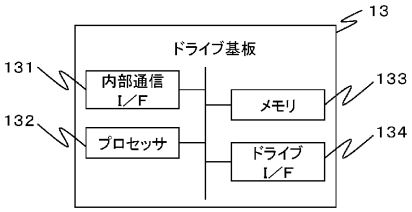
【図５】



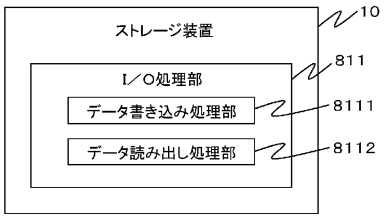
【図６】



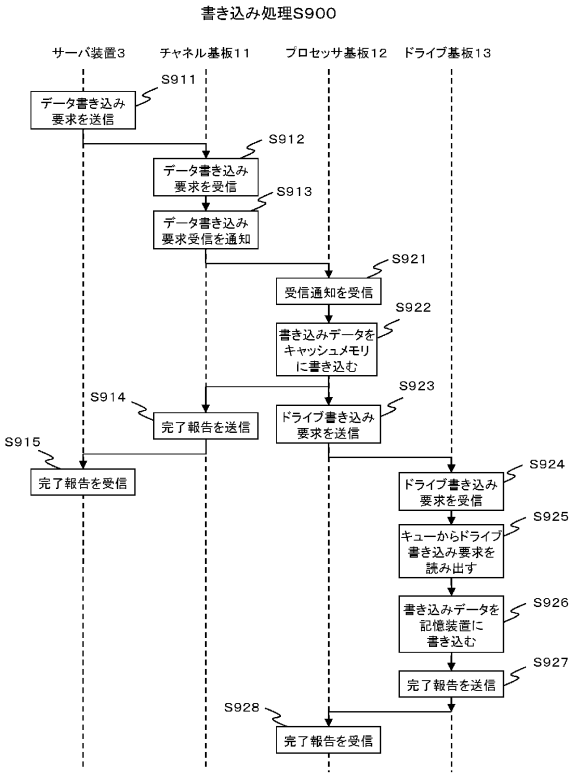
【図７】



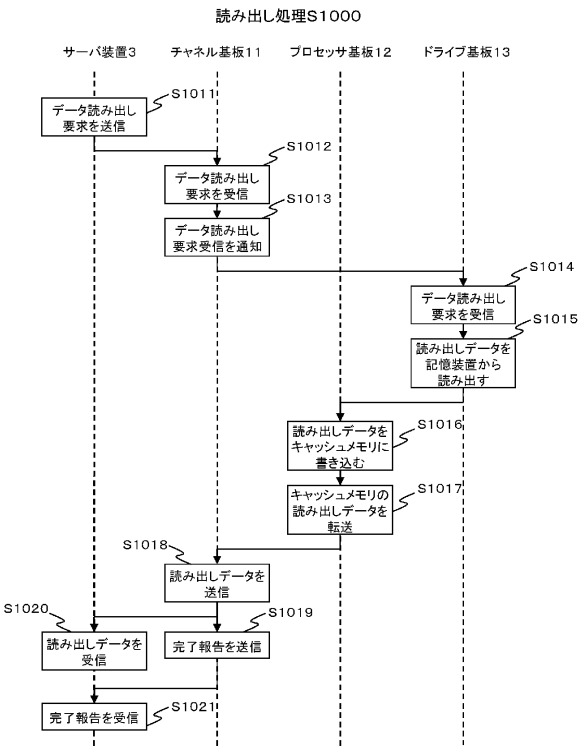
【図８】



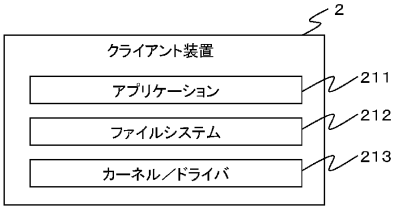
【図９】



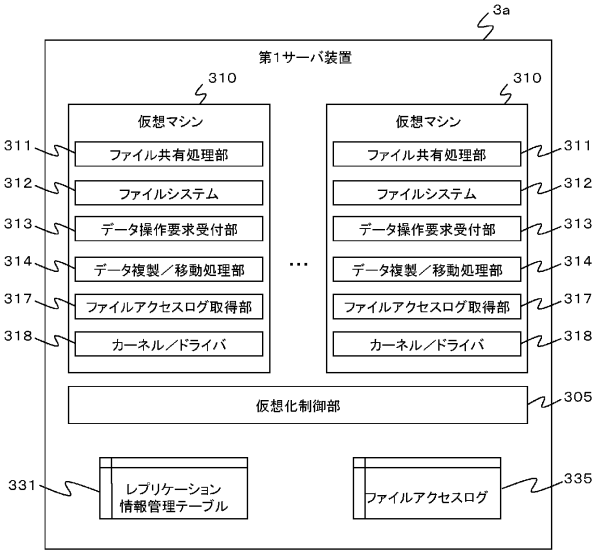
【図１０】



【図１１】



【図１２】



【図１３】

レプリケーション情報管理テーブル331

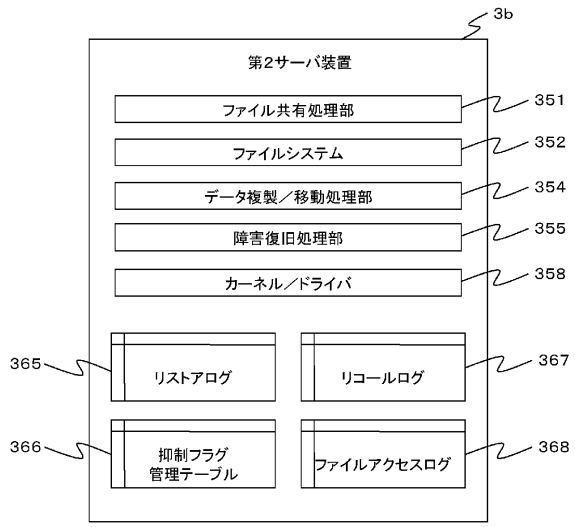
3311	レプリケーション先：host-001，201.123.45.6
3312	スタブ化閾値：200GB

【図１４】

ファイルアクセスログ 335

3351	3352	3353
2010/6/16 12:00:00	/home/user01/a.txt	user01
2010/6/16 12:10:00	/home/user02/b.txt	user02
2010/6/16 12:20:00	/home/user01/a.txt	user03
：	：	：

【図１５】



【図１６】

リストアログ365

3651	3652	3653
日時	イベント	リストア対象ファイル
2010/9/4 00:00:00	リストア開始	
2010/9/4 00:05:00	リストア実行	/home/user01/a.txt
2010/9/4 00:07:00	リストア実行	/home/user02/a.txt
：	：	：

【図１７】

抑制フラグ管理テーブル366

3661	抑制フラグ：ON
3662	更新日時 ：2010／9／4 00:05:14

【図１８】

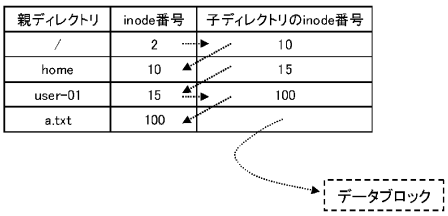
リコールログ367

3671	3672
日時	リコール対象ファイル
2010/9/4 00:05:00	/home/user01/a.txt
2010/9/4 00:07:00	/home/user02/a.txt
：	：

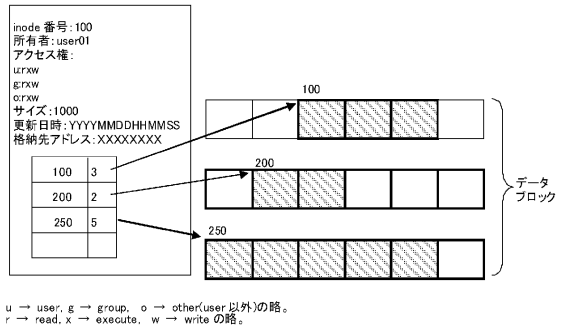
【図１９】

ファイルシステム構造1900

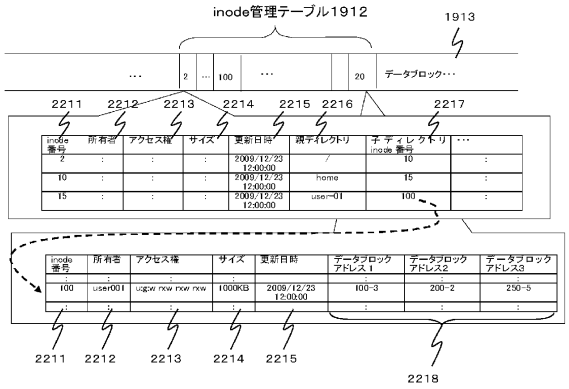
【図２０】



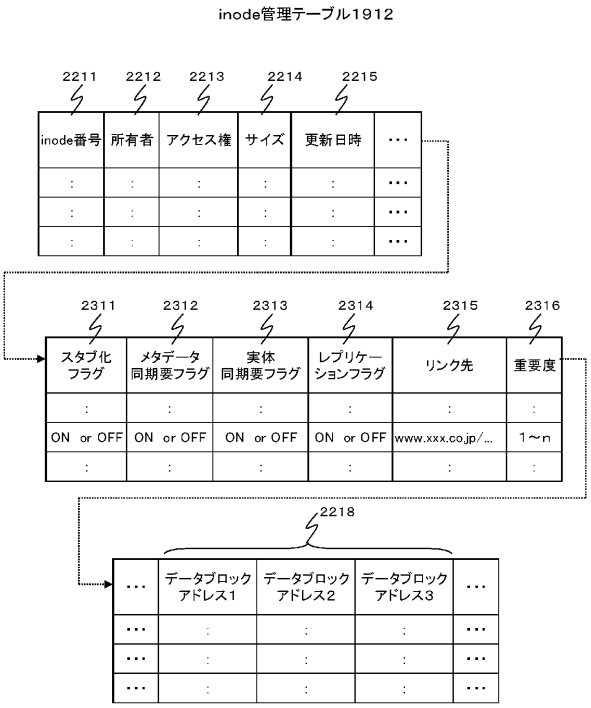
【図２１】



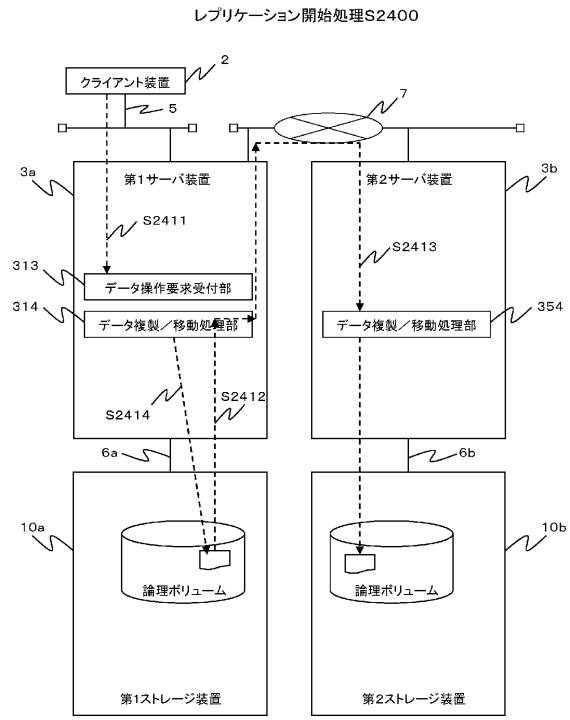
【図２２】



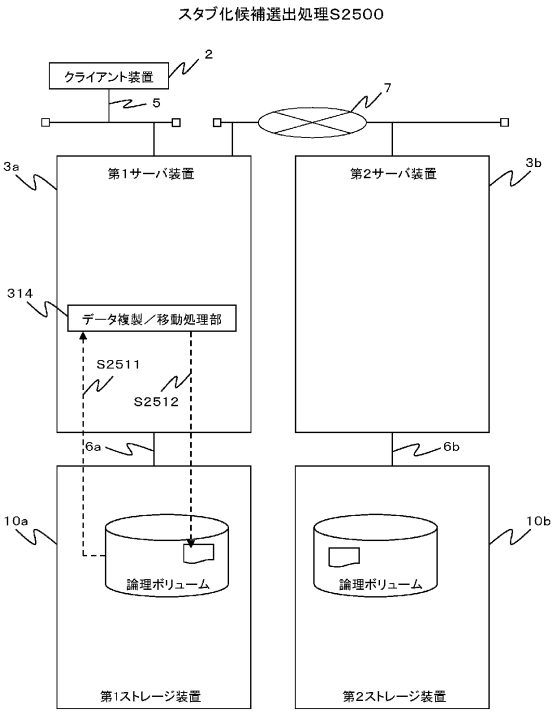
【図２３】



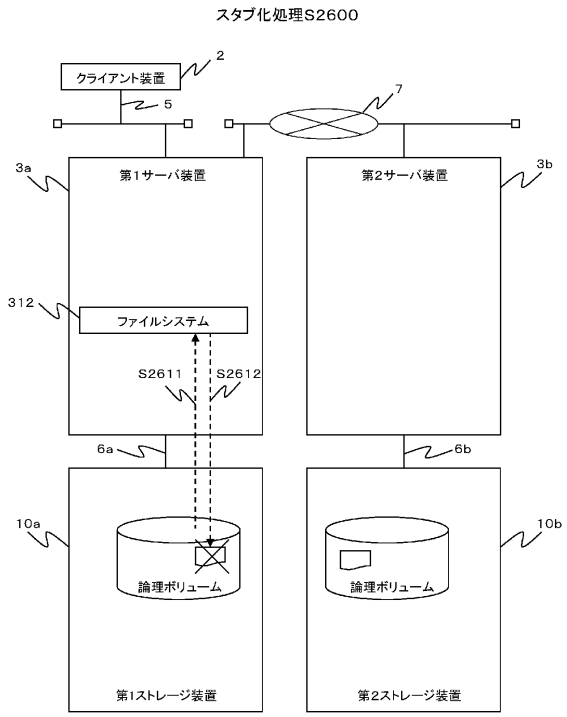
【図２４】



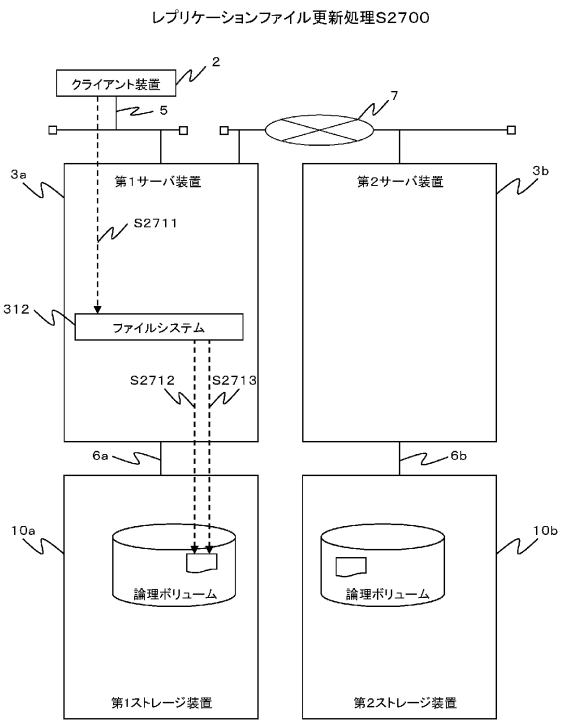
【図２５】



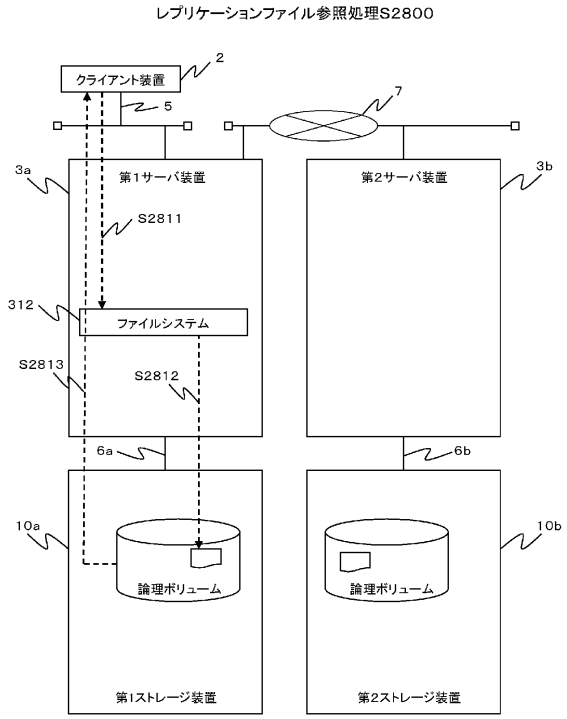
【図２６】



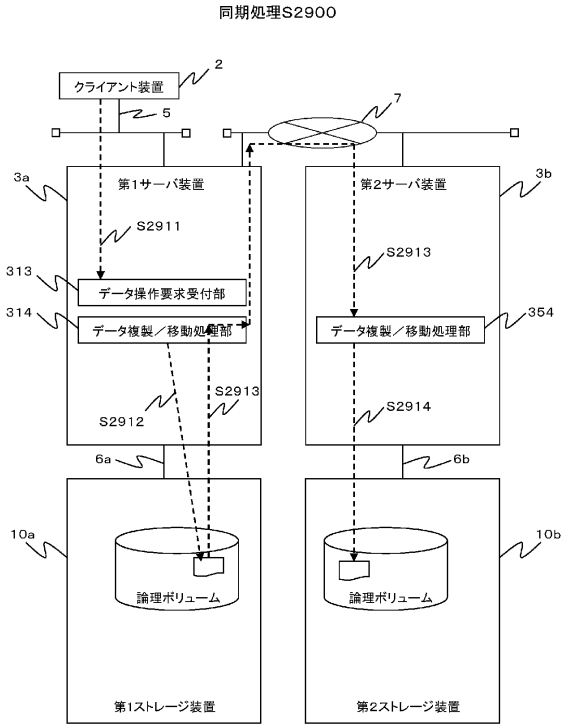
【図２７】



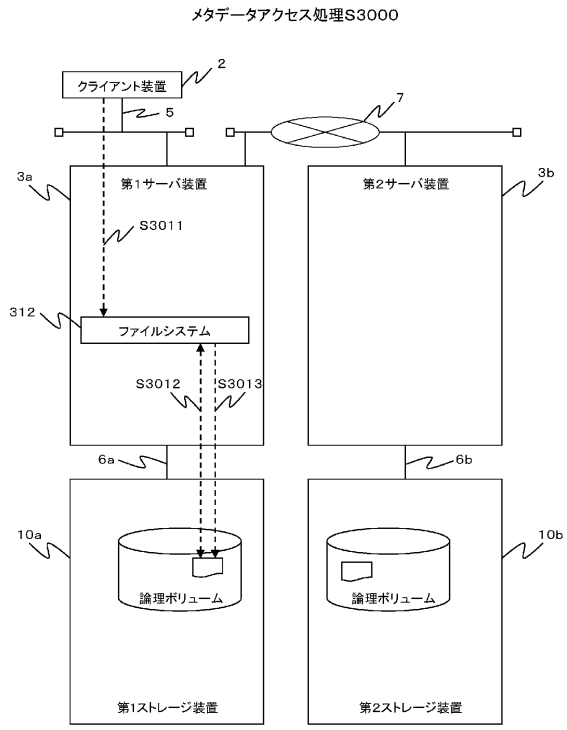
【図２８】



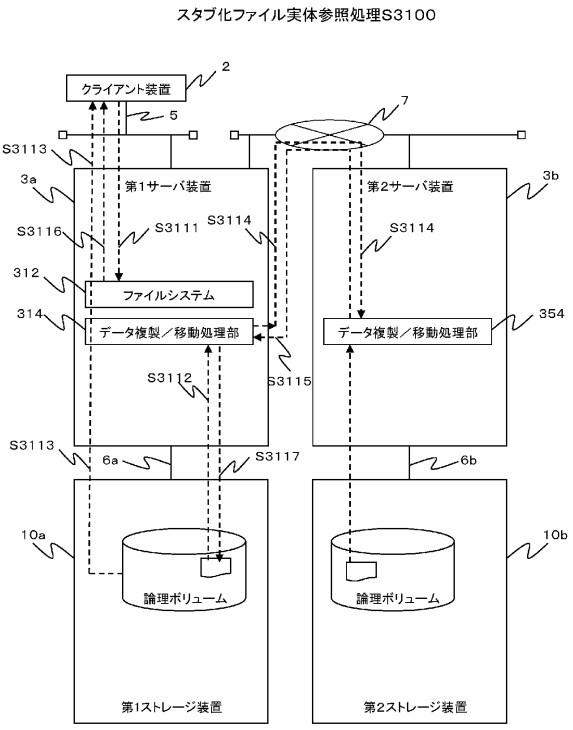
【図２９】



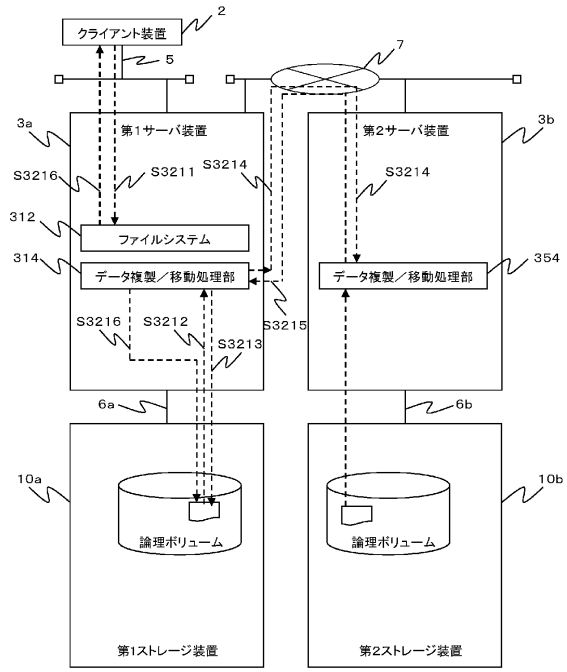
【図３０】



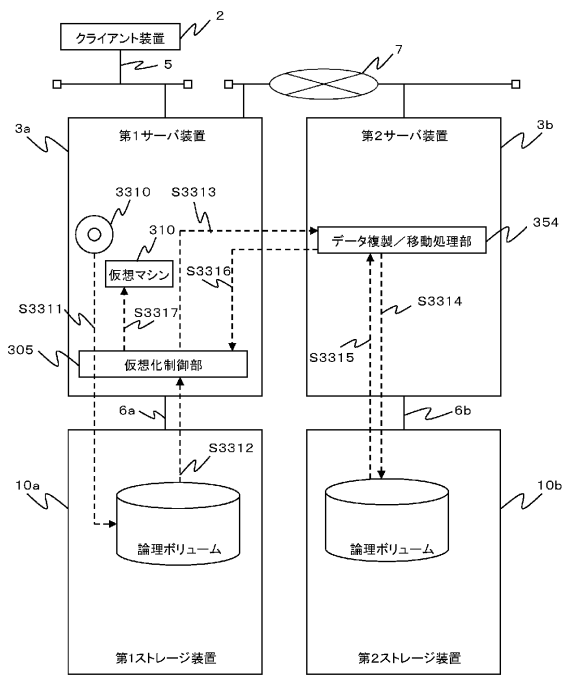
【図３１】



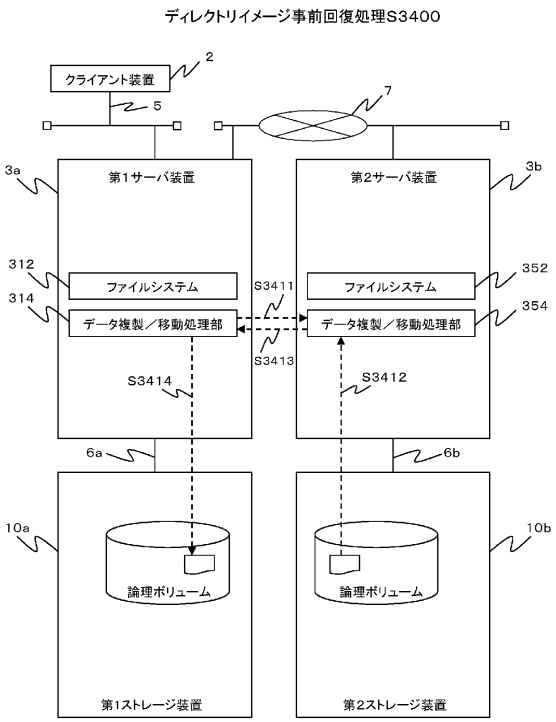
【図３２】



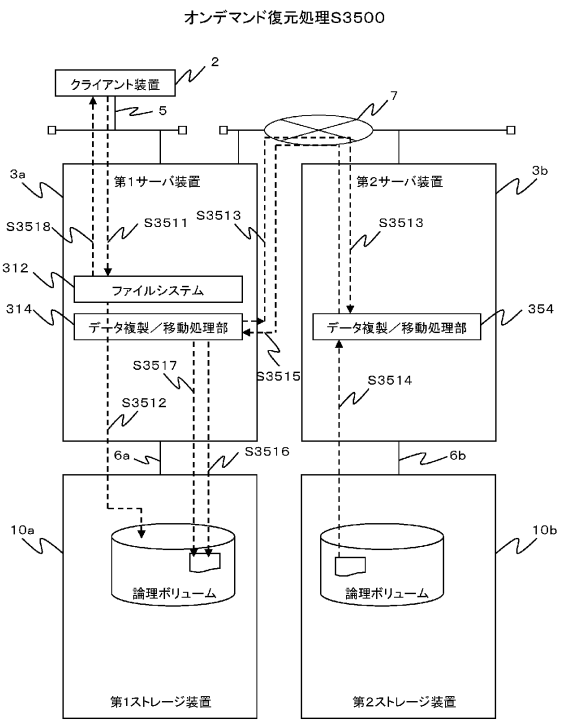
【図３３】



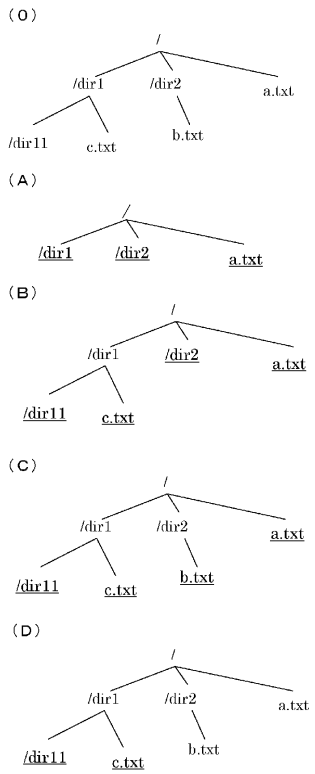
【図３４】



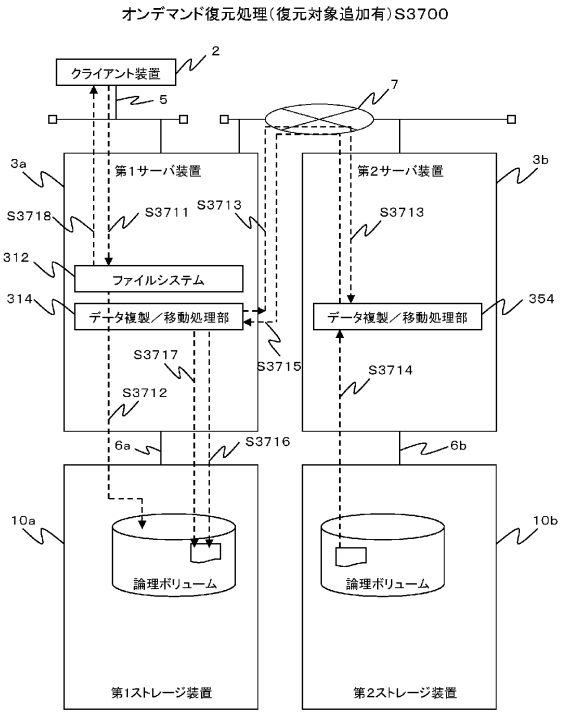
【図３５】



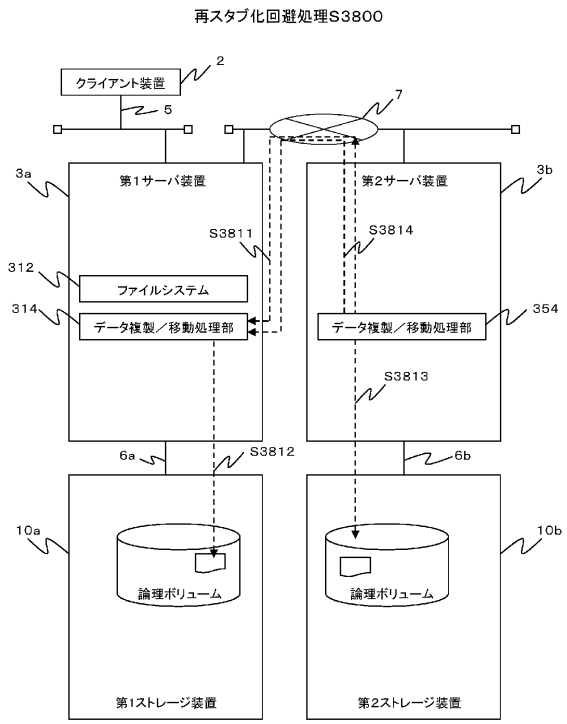
【図３６】



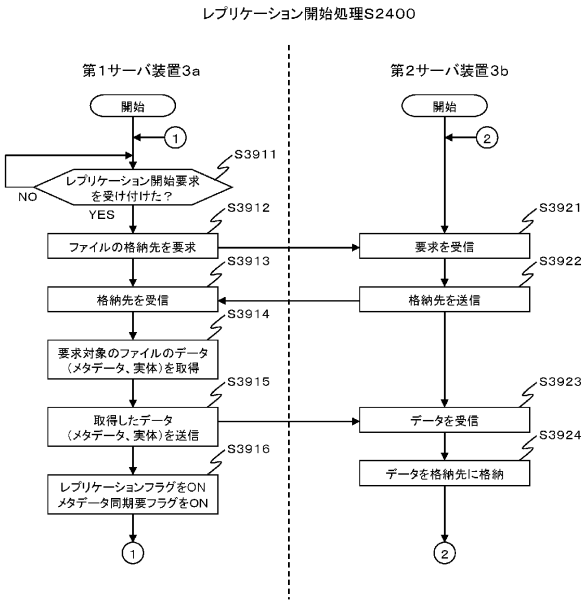
【図３７】



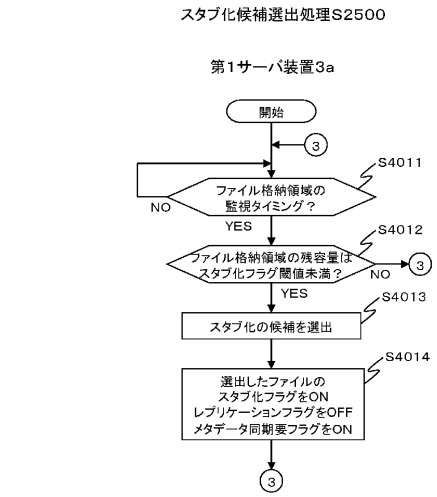
【図３８】



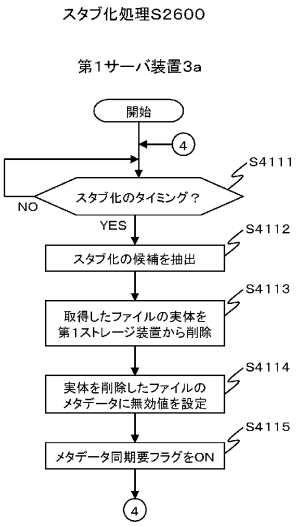
【図３９】



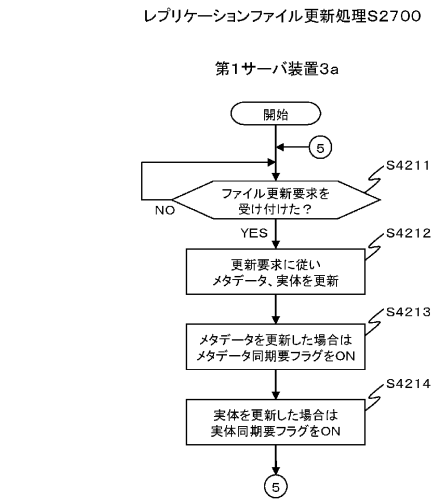
【図４０】



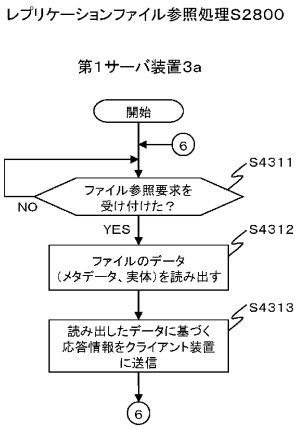
【図４１】



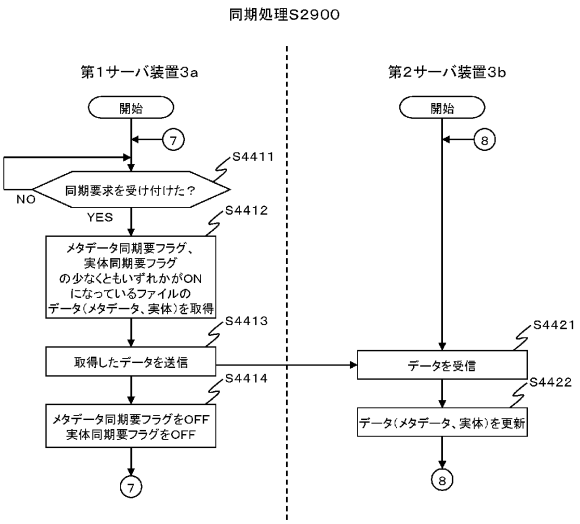
【図４２】



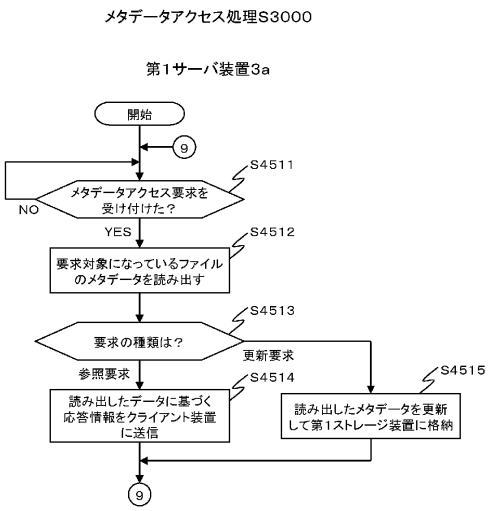
【図４３】



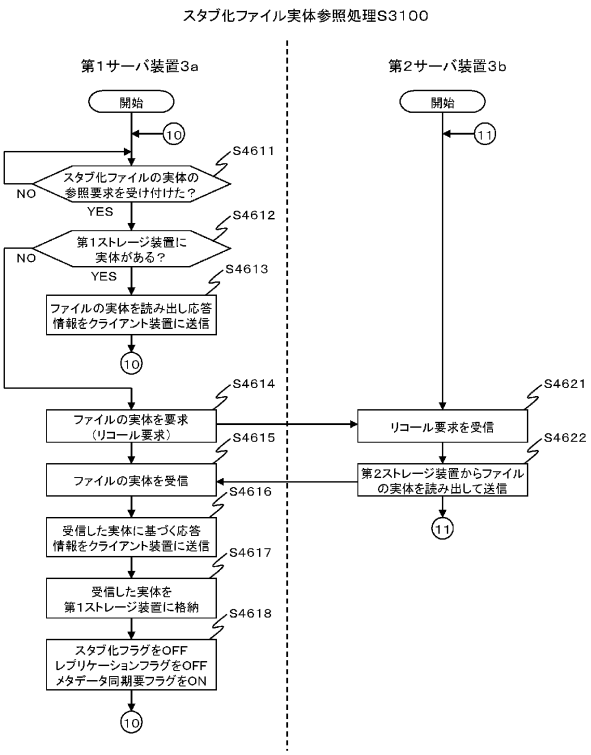
【図４４】



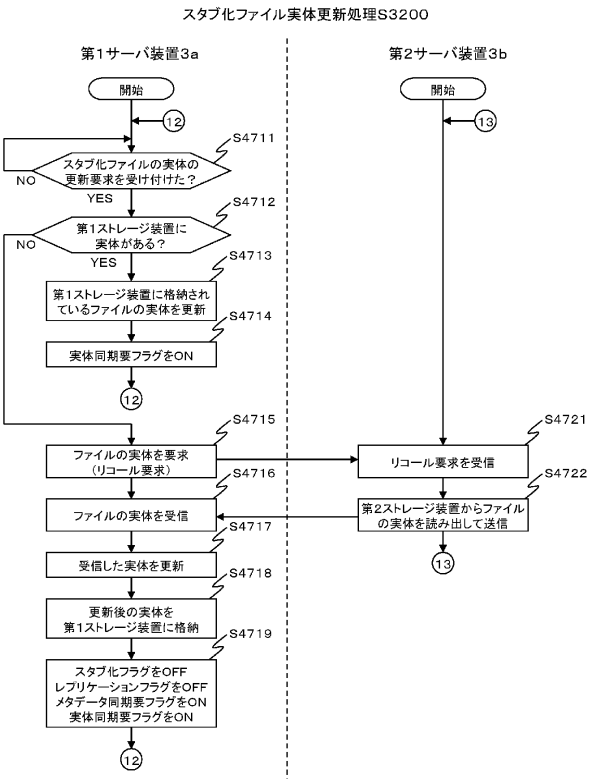
【図４５】



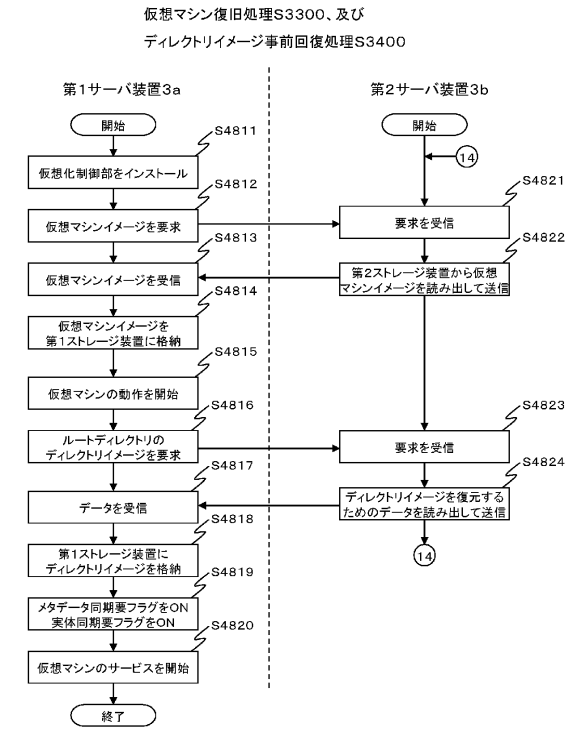
【図４６】



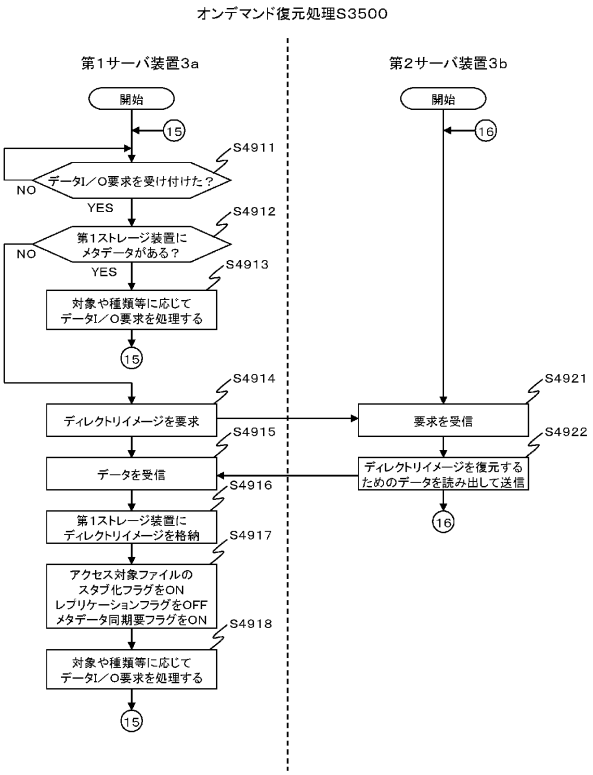
【図４７】



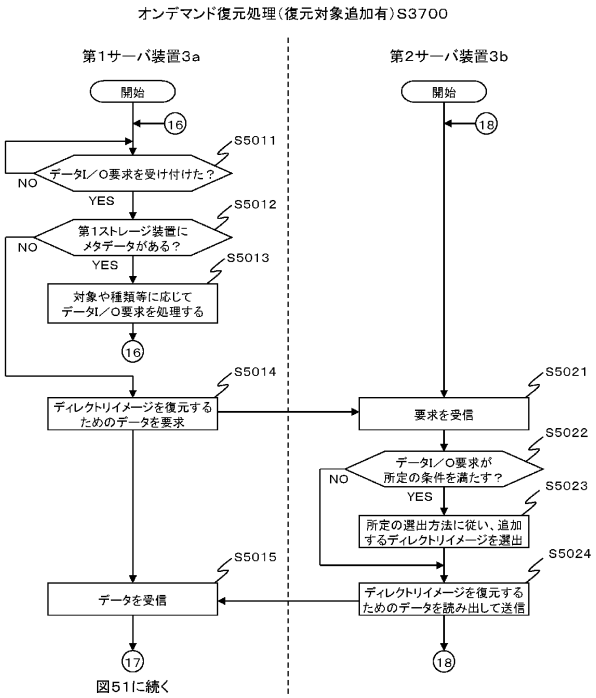
【図４８】



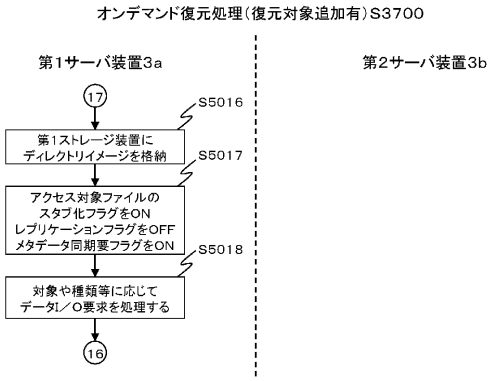
【図４９】



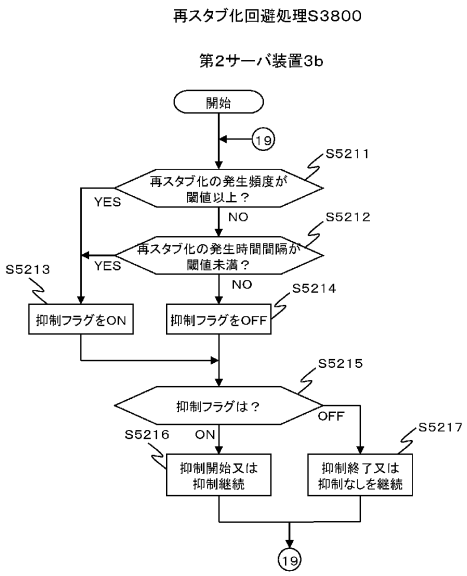
【図５０】


オンデマンド復元処理（復元対象追加有）S3700
第1サーバ装置3a
第2サーバ装置3b
開始
16
S5011
データI／O要求を受け付けた？
NO
YES
S5012
第1ストレージ装置にメタデータがある？
NO
YES
S5013
対象や種類等に応じてデータI／O要求を処理する
16
S5014
ディレクトリイメージを復元するためのデータを要求
S5015
データを受信
17
図51に続く
開始
18
S5021
要求を受信
S5022
データI／O要求が所定の条件を満たす？
NO
YES
S5023
所定の選出方法に従い、追加するディレクトリイメージを選出
S5024
ディレクトリイメージを復元するためのデータを読み出して送信
18

【図５１】


オンデマンド復元処理（復元対象追加有）S3700
第1サーバ装置3a
第2サーバ装置3b
17
S5016
第1ストレージ装置にディレクトリイメージを格納
S5017
アクセス対象ファイルのスタブ化フラグをON
レプリケーションフラグをOFF
メタデータ同期要フラグをON
S5018
対象や種類等に応じてデータI／O要求を処理する
16

【図５２】


再スタブ化回避処理S3800
第2サーバ装置3b
開始
19
S5211
再スタブ化の発生頻度が閾値以上？
YES
NO
S5212
再スタブ化の発生時間間隔が閾値未満？
YES
NO
S5213
抑制フラグをON
S5214
抑制フラグをOFF
S5215
抑制フラグは？
ON
OFF
S5216
抑制開始又は抑制継続
S5217
抑制終了又は抑制なしを継続
19

フロントページの続き

(56)参考文献 特開２００５－５５９４７（ＪＰ，Ａ）
特開２００５－１４１５５５（ＪＰ，Ａ）
特表２００６－５０８４７３（ＪＰ，Ａ）
特開２００８－３３５１９（ＪＰ，Ａ）
特開２００７－２８００９９（ＪＰ，Ａ）
米国特許出願公開第２００３／０１５８８６２（ＵＳ，Ａ１）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　１２／００
Ｇ０６Ｆ　３／０６