(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4472995号
(P4472995)
(45)発行日　平成22年6月2日(2010.6.2)　(24)登録日　平成22年3月12日(2010.3.12)

(51)Int.Cl.　F I
G06F 12/00 (2006.01)　G06F 12/00 531D
G06F 3/06 (2006.01)　G06F 3/06 304F

請求項の数 18　(全 21 頁)

(21)出願番号　特願2003-575277 (P2003-575277)
(86)(22)出願日　平成15年3月6日(2003.3.6)
(65)公表番号　特表2005-519408 (P2005-519408A)
(43)公表日　平成17年6月30日(2005.6.30)
(86)国際出願番号　PCT/US2003/006620
(87)国際公開番号　WO2003/077128
(87)国際公開日　平成15年9月18日(2003.9.18)
審査請求日　平成18年3月6日(2006.3.6)
(31)優先権主張番号　10/090,728
(32)優先日　平成14年3月6日(2002.3.6)
(33)優先権主張国　米国(US)

(73)特許権者　500020944
マラソン　テクノロジーズ　コーポレイション
アメリカ合衆国　マサチューセッツ州　01719　ボックスボロ　マサチューセッツ　アベニュー　1300
(74)代理人　100082005
弁理士　熊倉　禎男
(74)代理人　100067013
弁理士　大塚　文昭
(74)代理人　100074228
弁理士　今城　俊夫
(74)代理人　100086771
弁理士　西島　孝喜

最終頁に続く

(54)【発明の名称】増分不一致を使用するミラーコピーの生成方法

(57)【特許請求の範囲】
【請求項１】
コンピュータシステムにおいて、第１のコントローラが揮発性記憶装置及び不揮発性記憶装置を具備する第１の記憶装置と関連し、第２のコントローラが揮発性記憶装置及び不揮発性記憶装置を具備する第２の記憶装置と関連しており、該第１の記憶装置のミラーコピーを該第２の記憶装置に保持する方法であって、
前記第１のコントローラが前記第１の記憶装置において書き込み要求を受け取る段階と、
前記第１の記憶装置の揮発性記憶装置又は不揮発性記憶装置に前記書き込み要求を保存するとともに、前記第１の記憶装置の揮発性記憶装置に保存されている書き込み要求を前記第１の記憶装置の不揮発性記憶装置に対して周期的にコミットすることによって、前記第１の記憶装置で受け取られた前記書き込み要求を前記第１のコントローラが処理する段階と、
前記第２の記憶装置の揮発性記憶装置又は不揮発性記憶装置に前記書き込み要求を保存するとともに、前記第２の記憶装置の揮発性記憶装置に保存されている書き込み要求を前記第２の記憶装置の不揮発性記憶装置に対して周期的にコミットすることによって、前記第２の記憶装置において書き込み要求を前記第２のコントローラが受け取る段階と、
前記第１の記憶装置によって処理された書き込み要求を識別する基準ラベルに基づく情報である書き込み要求を指定する情報と共に同期コミットメッセージを前記第１のコントローラが前記第２の記憶装置に周期的に送る段階と、

前記第２のコントローラが、前記第２の記憶装置の揮発性記憶装置又は不揮発性記憶装置を用いて、前記同期コミットメッセージにおいて識別された前記書き込み要求を処理した後で、前記第２のコントローラが揮発性ディスク記憶装置において保存されている書き込み要求を前記第２の記憶装置の不揮発性ディスク記憶装置にコミットする段階と、

前記同期コミットメッセージを受け取った後で、前記処理された書き込み要求が前記第２の記憶装置の不揮発性記憶装置に正常にコミットされた場合に、前記指定された書き込み要求に先行していたすべての書き込み要求に関連する不揮発性記憶装置書き込み対象データが前記第２の記憶装置の不揮発性記憶装置に書き込まれたことと、前記指定された書き込み要求に関連する不揮発性記憶装置書き込み対象データが前記第２の記憶装置の不揮発性記憶装置に書き込まれたこととを示す情報を含む確認メッセージを前記第２の記憶装置によって前記第１のコントローラに送る段階と、を含む方法。

【請求項２】

前記同期コミットメッセージを受け取った後で、前記第２の記憶装置が前記指定された書き込み要求を処理する段階を更に含む請求項１に記載の方法。

【請求項３】

コンピュータシステムにおいて、第１のコントローラが揮発性記憶装置及び不揮発性記憶装置を具備する第１の記憶装置と関連し、第２のコントローラが揮発性記憶装置及び不揮発性記憶装置を具備する第２の記憶装置と関連しており、該第１の記憶装置のミラーコピーを該第２の記憶装置に保持する方法であって、

前記第１のコントローラが前記第１の記憶装置において書き込み要求を受け取る段階と、

前記第１の記憶装置の揮発性記憶装置又は不揮発性記憶装置に前記書き込み要求を保存するとともに、前記第１の記憶装置の揮発性記憶装置に保存されている書き込み要求を前記第１の記憶装置の不揮発性記憶装置に対して周期的にコミットすることによって、前記第１の記憶装置で受け取られた前記書き込み要求を前記第１のコントローラが処理する段階と、

前記第２の記憶装置の揮発性記憶装置又は不揮発性記憶装置に前記書き込み要求を保存するとともに、前記第２の記憶装置の揮発性記憶装置に保存されている書き込み要求を前記第２の記憶装置の不揮発性記憶装置に対して周期的にコミットすることによって、前記第２の記憶装置において書き込み要求を前記第２のコントローラが受け取る段階と、

前記第２の記憶装置の揮発性記憶装置又は不揮発性記憶装置に前記書き込み要求を保存するとともに、前記第２の記憶装置の揮発性記憶装置に保存されている書き込み要求を前記第２の記憶装置の不揮発性記憶装置に対して周期的にコミットすることによって、前記第２の記憶装置において受け取られた前記書き込み要求を前記第２のコントローラが処理する段階と、

前記第１の記憶装置によって処理された書き込み要求を識別する基準ラベルに基づく情報である書き込み要求を指定する情報と共に同期コミットメッセージを前記第１のコントローラが前記第２の記憶装置に周期的に送る段階と、

前記第２のコントローラが、前記第２の記憶装置の揮発性記憶装置又は不揮発性記憶装置を用いて、前記同期コミットメッセージにおいて識別された前記書き込み要求を処理した後で、前記第２のコントローラが揮発性ディスク記憶装置において保存されている書き込み要求を前記第２の記憶装置の不揮発性ディスク記憶装置にコミットする段階と、

前記同期コミットメッセージを受け取った後で、前記処理された書き込み要求が前記第２の記憶装置の不揮発性記憶装置に正常にコミットされた場合に、前記第２の記憶装置の揮発性記憶装置の正常なキャッシュフラッシュを示す情報を含む確認メッセージを前記第２の記憶装置によって前記第１のコントローラに送る段階と、を含む方法。

【請求項４】

コンピュータシステムにおいて、第１のコントローラが揮発性記憶装置及び不揮発性記憶装置を具備する第１の記憶装置と関連し、第２のコントローラが揮発性記憶装置及び不揮発性記憶装置を具備する第２の記憶装置と関連しており、該第１の記憶装置のミラーコ

ピーを該第２の記憶装置に保持する方法であって、
前記第１のコントローラが前記第１の記憶装置において書き込み要求を受け取る段階と、
前記第１の記憶装置の揮発性記憶装置又は不揮発性記憶装置に前記書き込み要求を保存するとともに、前記第１の記憶装置の揮発性記憶装置に保存されている書き込み要求を前記第１の記憶装置の不揮発性記憶装置に対して周期的にコミットすることによって、前記第１の記憶装置で受け取られた前記書き込み要求を前記第１のコントローラが処理する段階と、
前記第２の記憶装置の揮発性記憶装置又は不揮発性記憶装置に前記書き込み要求を保存するとともに、前記第２の記憶装置の揮発性記憶装置に保存されている書き込み要求を前記第２の記憶装置の不揮発性記憶装置に対して周期的にコミットすることによって、前記第２の記憶装置において書き込み要求を前記第２のコントローラが受け取る段階と、
前記第１の記憶装置によって処理された書き込み要求を識別する基準ラベルであって、書き込み要求に対して順次に割り当てられる基準ラベルに基づく情報である書き込み要求を指定する情報と共に同期コミットメッセージを前記第１のコントローラが前記第２の記憶装置に周期的に送る段階と、
前記第２のコントローラが、前記第２の記憶装置の揮発性記憶装置又は不揮発性記憶装置を用いて、前記同期コミットメッセージにおいて識別された前記書き込み要求を処理した後で、前記第２のコントローラが揮発性ディスク記憶装置において保存されている書き込み要求を前記第２の記憶装置の不揮発性ディスク記憶装置にコミットする段階と、
前記同期コミットメッセージを受け取った後で、前記処理された書き込み要求が前記第２の記憶装置の不揮発性記憶装置に正常にコミットされた場合に、前記指定された書き込み要求に関連するデータが前記第２の記憶装置の不揮発性記憶装置に書き込まれたことを示す情報を含む確認メッセージを前記第２の記憶装置によって前記第１のコントローラに送る段階と、を含む方法。
【請求項５】
前記第２の記憶装置の揮発性記憶装置又は不揮発性記憶装置に書き込み要求が保存される前に、前記同期コミットメッセージの基準ラベルによって識別された該書き込み要求に先立って発行された全ての書き込み要求を、前記第２の記憶装置の揮発性記憶装置又は不揮発性記憶装置に保存するように、前記第２のコントローラが書き込み要求を前記第２の記憶装置において受け取る段階は、前記第２のコントローラが要求をこの要求の基準ラベルによって順次に書き込む段階を含むことを特徴とする請求項４に記載の方法。
【請求項６】
各記憶装置が基準ラベルと同じ順番で書き込み要求を受け取ることを特徴とする請求項４に記載の方法。
【請求項７】
前記第１の記憶装置の揮発性記憶装置又は不揮発性記憶装置に保存された書き込み要求によって影響を受ける記憶装置の領域を前記第１のコントローラが識別する段階を更に含む請求項４に記載の方法。
【請求項８】
前記第１の記憶装置の揮発性記憶装置又は不揮発性記憶装置において保存されていた書き込み要求によって影響を受ける記憶装置の領域を前記第１のコントローラが識別する段階は、第１のビットマップ内に前記識別された記憶装置の領域を前記第１のコントローラが蓄積する段階を更に含む請求項７に記載の方法。
【請求項９】
コンピュータシステムにおいて、第１のコントローラが揮発性記憶装置及び不揮発性記憶装置を具備する第１の記憶装置と関連し、第２のコントローラが揮発性記憶装置及び不揮発性記憶装置を具備する第２の記憶装置と関連しており、該第１の記憶装置のミラーコピーを該第２の記憶装置に保持する方法であって、
前記第１の記憶装置の揮発性記憶装置又は不揮発性記憶装置に前記書き込み要求を保存

するとともに、前記第１の記憶装置の揮発性記憶装置に保存されている書き込み要求を前記第１の記憶装置の不揮発性記憶装置に対して周期的にコミットすることによって、前記第１のコントローラが前記第１の記憶装置において書き込み要求を受け取る段階と、

前記第１の記憶装置の揮発性記憶装置又は不揮発性記憶装置に前記書き込み要求を保存するとともに、前記第１の記憶装置の揮発性記憶装置に保存されている書き込み要求を前記第１の記憶装置の不揮発性記憶装置に対して周期的にコミットすることによって、前記第１の記憶装置で受け取られた前記書き込み要求を前記第１のコントローラが処理する段階と、

前記第２の記憶装置の揮発性記憶装置又は不揮発性記憶装置に前記書き込み要求を保存するとともに、前記第２の記憶装置の揮発性記憶装置に保存されている書き込み要求を前記第２の記憶装置の不揮発性記憶装置に対して周期的にコミットすることによって、前記第２の記憶装置において書き込み要求を前記第２のコントローラが受け取る段階と、

前記第１の記憶装置によって処理された書き込み要求を識別する基準ラベルに基づく情報である書き込み要求を指定する情報と共に同期コミットメッセージを前記第１のコントローラが前記第２の記憶装置に周期的に送る段階と、

前記第２のコントローラが、前記第２の記憶装置の揮発性記憶装置又は不揮発性記憶装置を用いて、前記同期コミットメッセージにおいて識別された前記書き込み要求を処理した後で、前記第２のコントローラが揮発性ディスク記憶装置において保存されている書き込み要求を前記第２の記憶装置の不揮発性ディスク記憶装置にコミットする段階と、

前記同期コミットメッセージを受け取った後で、前記処理された書き込み要求が前記第２の記憶装置の不揮発性記憶装置に正常にコミットされた場合に、前記指定された書き込み要求に関連する不揮発性記憶装置書き込み対象データが前記第２の記憶装置の不揮発性記憶装置に書き込まれたことを示す情報を含む確認メッセージを前記第２の記憶装置によって前記第１のコントローラに送る段階と、

前記第１の記憶装置の揮発性記憶装置又は不揮発性記憶装置に保存された書き込み要求によって影響を受ける記憶装置の領域を前記第１のコントローラが識別する段階と、

前記第１のコントローラが前記識別された領域を第１のビットマップ中に蓄積する段階と、

前記同期コミットメッセージを送った後で、第２のビットマップ内に新規に識別された記憶装置の領域を前記第１のコントローラが蓄積する段階と、

前記処理された書き込み要求のデータが前記第２の記憶装置の不揮発性記憶装置に書き込まれたことを前記第２の記憶装置が確認した後で、前記書き込みデータが前記不揮発性記憶装置に正常に書き込まれたかどうかを示すステータスメッセージを前記第２のコントローラによって前記第１の記憶装置に送る段階と、

前記書き込みデータが正常に書き込まれたことを示す前記ステータスメッセージを受け取った後で、前記第１のビットマップを前記第１のコントローラが削除する段階と、を含む方法。

【請求項１０】

前記書き込みデータが正常に書き込まれなかったことを示すステータスメッセージを前記第１のコントローラが受け取った後で、前記第２のビットマップの内容を前記第１のビットマップに前記第１のコントローラがコピーし、前記第２のビットマップを前記第１のコントローラが削除する段階を更に含む請求項９に記載の方法。

【請求項１１】

前記第１のビットマップを削除した後で、前記第２のビットマップを前記第１のビットマップとして前記第１のコントローラが指定する段階を更に含む請求項９に記載の方法。

【請求項１２】

コンピュータシステムにおいて、第１のコントローラが揮発性記憶装置及び不揮発性記憶装置を具備する第１の記憶装置と関連し、第２のコントローラが揮発性記憶装置及び不揮発性記憶装置を具備する第２の記憶装置と関連しており、該第１の記憶装置のミラーコピーを該第２の記憶装置に保持する方法であって、

前記第１のコントローラが前記第１の記憶装置において書き込み要求を受け取る段階と、

前記第１の記憶装置の揮発性記憶装置又は不揮発性記憶装置に前記書き込み要求を保存するとともに、前記第１の記憶装置の揮発性記憶装置に保存されている書き込み要求を前記第１の記憶装置の不揮発性記憶装置に対して周期的にコミットすることによって、前記第１の記憶装置で受け取られた前記書き込み要求を前記第１のコントローラが処理する段階と、

前記第２の記憶装置の揮発性記憶装置又は不揮発性記憶装置に前記書き込み要求を保存するとともに、前記第２の記憶装置の揮発性記憶装置に保存されている書き込み要求を前記第２の記憶装置の不揮発性記憶装置に対して周期的にコミットすることによって、前記第２の記憶装置において書き込み要求を前記第２のコントローラが受け取る段階と、

前記第１の記憶装置によって処理された書き込み要求を識別する基準ラベルに基づく情報である書き込み要求を指定する情報と共に同期コミットメッセージを前記第１のコントローラが前記第２の記憶装置に周期的に送る段階と、

前記第２のコントローラが、前記第２の記憶装置の揮発性記憶装置又は不揮発性記憶装置を用いて、前記同期コミットメッセージにおいて識別された前記書き込み要求を処理した後で、前記第２のコントローラが揮発性ディスク記憶装置において保存されている書き込み要求を前記第２の記憶装置の不揮発性ディスク記憶装置にコミットする段階と、

前記同期コミットメッセージを受け取った後で、前記処理された書き込み要求が前記第２の記憶装置の不揮発性記憶装置に正常にコミットされた場合に、前記指定された書き込み要求に関連する不揮発性記憶装置書き込み対象データが前記第２の記憶装置の不揮発性記憶装置に書き込まれたことを示す情報を含む確認メッセージを前記第２の記憶装置によって前記第１のコントローラに送る段階と、

前記第２の記憶装置が書き込み要求を処理することができなかった期間の後で、

前記第１のコントローラが前記第１のビットマップの内容を回復ビットマップにコピーする段階と、

前記回復ビットマップを使用して、前記第１の記憶装置から前記第２の記憶装置にコピーされることになる前記第１の記憶装置の記憶領域を前記第１のコントローラが識別する段階と、

前記第１の記憶装置の識別された記憶領域を前記第１のコントローラが前記第２の記憶装置にコピーする段階と、

第３のビットマップ内に前記第１の記憶装置において新規に受け取られた書き込み要求を前記第１のコントローラが蓄積する段階と、

を含む方法。

【請求項１３】

コンピュータシステムにおいて、第１のコントローラが揮発性記憶装置及び不揮発性記憶装置を具備する第１の記憶装置と関連し、第２のコントローラが揮発性記憶装置及び不揮発性記憶装置を具備する第２の記憶装置と関連しており、該第１の記憶装置のミラーコピーを該第２の記憶装置に保持する方法であって、

前記第１のコントローラが前記第１の記憶装置において書き込み要求を受け取る段階と、

前記第１の記憶装置の揮発性記憶装置又は不揮発性記憶装置に前記書き込み要求を保存するとともに、前記第１の記憶装置の揮発性記憶装置に保存されている書き込み要求を前記第１の記憶装置の不揮発性記憶装置に対して周期的にコミットすることによって、前記第１の記憶装置で受け取られた前記書き込み要求を前記第１のコントローラが処理する段階と、

前記第２の記憶装置の揮発性記憶装置又は不揮発性記憶装置に前記書き込み要求を保存するとともに、前記第２の記憶装置の揮発性記憶装置に保存されている書き込み要求を前記第２の記憶装置の不揮発性記憶装置に対して周期的にコミットすることによって、前記第２の記憶装置において書き込み要求を前記第２のコントローラが受け取る段階と、

前記第１の記憶装置によって処理された書き込み要求を識別する基準ラベルに基づく情報である書き込み要求を指定する情報と共に同期コミットメッセージを前記第１のコントローラが前記第２の記憶装置に周期的に送る段階と、
前記同期コミットメッセージを受け取った後で、前記処理された書き込み要求が前記第２の記憶装置の不揮発性記憶装置に正常にコミットされた場合に、前記指定された書き込み要求に関連する不揮発性記憶装置書き込み対象データが前記第２の記憶装置の不揮発性記憶装置に書き込まれたことを示す情報を含む確認メッセージを前記第２の記憶装置によって前記第１のコントローラに送る段階と、
前記第２の記憶装置によって処理された書き込み要求を識別する基準ラベルに基づく情報である第２の書き込み要求を指定する情報と共に第２の同期コミットメッセージを前記第２のコントローラが前記第１の記憶装置に送る段階と、
前記第２の同期コミットメッセージを受け取った後で、前記処理された書き込み要求が前記第１の記憶装置の不揮発性記憶装置に正常にコミットされた場合に、前記指定された第２の書き込み要求に関連する不揮発性記憶装置書き込み対象データが前記第１の記憶装置の不揮発性記憶装置に書き込まれたことを示す情報を含む確認メッセージを前記第１の記憶装置によって前記第２のコントローラに送る段階と、
を含む方法。
【請求項１４】
前記処理された書き込み要求が前記第１の記憶装置の不揮発性記憶装置に正常にコミットされた場合に、前記指定された第２の書き込み要求に関連するデータが前記第１の記憶装置の不揮発性記憶装置に書き込まれたことを示す情報を含む確認メッセージを前記第１の記憶装置が前記第２のコントローラに送る段階は、前記指定された第２の書き込み要求に先行する全ての書き込み要求に関連するデータが前記第１の記憶装置の不揮発性記憶装置に書き込まれているデータである確認メッセージを前記第１の記憶装置が前記第２のコントローラに送る段階を含む請求項１３に記載の方法。
【請求項１５】
コンピュータシステムにおいて、第１のコントローラが揮発性記憶装置及び不揮発性記憶装置を具備する第１の記憶装置と関連し、第２のコントローラが揮発性記憶装置及び不揮発性記憶装置を具備する第２の記憶装置と関連しており、該第１の記憶装置のミラーコピーを該第２の記憶装置に保持する方法であって、
前記第１のコントローラが前記第１の記憶装置において書き込み要求を受け取る段階と、
前記第１の記憶装置の揮発性記憶装置又は不揮発性記憶装置に前記書き込み要求を保存するとともに、前記第１の記憶装置の揮発性記憶装置に保存されている書き込み要求を前記第１の記憶装置の不揮発性記憶装置に対して周期的にコミットすることによって、前記第１の記憶装置で受け取られた前記書き込み要求を前記第１のコントローラが処理する段階と、
前記第２の記憶装置の揮発性記憶装置又は不揮発性記憶装置に前記書き込み要求を保存するとともに、前記第２の記憶装置の揮発性記憶装置に保存されている書き込み要求を前記第２の記憶装置の不揮発性記憶装置に対して周期的にコミットすることによって、前記第２の記憶装置において書き込み要求を前記第２のコントローラが受け取る段階と、
前記第１の記憶装置によって処理された書き込み要求を識別する基準ラベルに基づく情報である書き込み要求を指定する情報と共に同期コミットメッセージを前記第１のコントローラが前記第２の記憶装置に周期的に送る段階と、
前記第２のコントローラが、前記第２の記憶装置の揮発性記憶装置又は不揮発性記憶装置を用いて、前記同期コミットメッセージにおいて識別された前記書き込み要求を処理した後で、前記第２のコントローラが揮発性ディスク記憶装置において保存されている書き込み要求を前記第２の記憶装置の不揮発性ディスク記憶装置にコミットする段階と、
前記同期コミットメッセージを受け取った後で、前記処理された書き込み要求が前記第２の記憶装置の不揮発性記憶装置に正常にコミットされた場合に、前記指定された書き込

み要求に関連するデータが前記第２の記憶装置の不揮発性記憶装置に書き込まれたことを示す情報を含む確認メッセージを前記第２の記憶装置によって前記第１のコントローラに送る段階と、
前記同期コミットメッセージを送った後で、ビットマップ内に新規の書き込み要求によって影響を受ける記憶装置の領域を前記第１のコントローラが蓄積する段階と、
前記第２の記憶装置が書き込み要求を処理することができるようになった後で、前記第１の記憶装置から前記第２の記憶装置へコピーされることになる前記第１の記憶装置の記憶領域を前記第１のコントローラが前記ビットマップを用いて識別する段階と、
前記第１の記憶装置の識別された領域の内容を前記第２のコントローラが前記第２の記憶装置にコピーする段階と、
を含む方法。
【請求項１６】
揮発性記憶装置及び不揮発性記憶装置を具備する第１の記憶装置と、
揮発性記憶装置及び不揮発性記憶装置を具備する第２の記憶装置と、
前記第１の記憶装置と関連付けられた第１のコントローラと、
前記第２の記憶装置と関連付けられた第２のコントローラと、
を備え、
前記第１のコントローラは、
前記第１の記憶装置において書き込み要求を受け取り、
前記第１の記憶装置の揮発性記憶装置又は不揮発性記憶装置に前記書き込み要求を保存するとともに、前記第１の記憶装置の揮発性記憶装置に保存されている書き込み要求を前記第１の記憶装置の不揮発性記憶装置に対して周期的にコミットすることによって、前記第１の記憶装置で受け取られた前記書き込み要求を処理し、
前記第１の記憶装置によって処理された書き込み要求を識別する基準ラベルに基づく情報である１つの書き込み要求を指定する情報と共に前記第２の記憶装置に対して同期コミットメッセージを周期的に送る、
ように構成されており、
前記第２のコントローラは、
前記第２の記憶装置において書き込み要求を受け取り、
前記第２のコントローラが、前記第２の記憶装置の揮発性記憶装置又は不揮発性記憶装置を用いて、前記同期コミットメッセージにおいて識別された前記書き込み要求を処理した後で、揮発性ディスク記憶装置において保存されている書き込み要求を前記第２の記憶装置の不揮発性ディスク記憶装置にコミットし、
前記同期コミットメッセージを受け取って該同期コミットメッセージ内の情報によって識別された前記書き込み要求を処理した後で、前記処理された書き込み要求が前記第２の記憶装置の不揮発性記憶装置に正常にコミットされた場合に、前記指定された書き込み要求に先行していたすべての書き込み要求に関連するデータが、前記第２の記憶装置の不揮発性記憶装置に書き込まれたことと前記指定された書き込み要求に関連するデータが前記第２の記憶装置の不揮発性記憶装置に書き込まれたこととを示す情報を含む確認メッセージを送るように構成されていることを特徴とするミラーデータ記憶システム。
【請求項１７】
前記第１のコントローラは、前記第１の記憶装置で処理された書き込み要求によって影響を受ける記憶装置の領域を識別するように更に構成されていることを特徴とする請求項１６に記載のシステム。
【請求項１８】
前記第１のコントローラは、第１のビットマップ内に前記記憶装置の識別された領域を蓄積するように構成されていることを特徴とする請求項１７に記載のシステム。
【発明の詳細な説明】
【技術分野】
【０００１】

本発明はディスクドライブ又は他の記憶装置のミラーコピーを生成するための手法に関する。

【背景技術】

【０００２】

多くのコンピュータシステムでは、同一のデータを複数の記憶装置の各々に記憶することによってフォールト・トレランスのレベルが設定される。同一のデータを有する記憶装置はミラー装置と呼ばれ、ミラーセットに属すると言われる。ミラーセットの１つのミラー装置が障害を発生するか又はアクセス不能となった場合、ミラーセットの他の１つ又は複数のミラー装置が継続してデータにアクセスする。

【０００３】

ミラーセットの各装置で同一のデータを保持するためには、各装置は、ミラーセットにデータを保存するために全ての要求（すなわち、全ての書き込み要求を）を受け取って処理しなければならない。ミラーセット内の装置は、該装置がこうした書き込み要求を処理できない場合には、ミラーセットの他の装置と不一致となる。ミラーセットのメンバーが不一致になると、１つのミラー装置から別のミラー装置にデータをコピーするためにミラーセットコピーを実行することができる。ミラーセットコピーを保持するための１つの手法においては、コンピュータシステムは停止され、１つのミラー装置から他のミラー装置に全てのデータがコピーされる。

【発明の開示】

【課題を解決するための手段】

【０００４】

１つの一般的な態様において、コンピュータシステムにおいて第１の記憶装置のミラーコピーが第２の記憶装置に保持される。第１の記憶装置は、関連するコントローラを含み、第２の記憶装置は、関連するコントローラと、揮発性記憶装置と、不揮発性記憶装置とを含む。記憶装置で受け取られた書き込み要求が処理される。同期コミットメッセージが、書き込み要求を指定する情報と共に第２の記憶装置に送られ、第２の記憶装置のコントローラが、同期コミットメッセージを受け取った後に、指定された書き込み要求に関連するデータが第２の記憶装置の不揮発性記憶装置に書き込まれたことを確認する。

【０００５】

実施形態は、１つ又はそれ以上の次の特徴を含む。例えば、第２の記憶装置のコントローラは、指定された書き込み要求に先行する全ての書き込み要求に関連するデータが第２の記憶装置の不揮発性記憶装置に書き込まれたことを確認することができる。或いは、第２の記憶装置のコントローラは、指定された書き込み要求を処理することができ、且つ指定された書き込み要求と先行する書き込み要求と関連するデータが第２の記憶装置の不揮発性記憶装置に書き込まれたことを確認することができる。第２の記憶装置のコントローラは、第２の記憶装置の揮発性記憶装置の正常なキャッシュフラッシュを確認することができる。

【０００６】

同期コミットメッセージと共に送られた情報は、第１の記憶装置によって処理されたか、或いは処理されることになる書き込み要求を識別する基準ラベルとすることができる。基準ラベルは、他の書き込み要求に割り当てられた基準ラベルに対して順次に割り当てることができる。第２の記憶装置で受け取られた全ての書き込み要求は、同期コミットメッセージの基準ラベルによって識別された書き込み要求の処理の前に順次処理することができる。

【０００７】

書き込み要求によって影響を受ける記憶装置の識別された領域は、例えば第１のビットマップに蓄積することができる。同期コミットメッセージを送った後は、記憶装置の新規に識別された領域は第２のビットマップに蓄積することができる。処理された書き込み要求のデータが第２の記憶装置の不揮発性記憶装置に書き込まれたことを、第２の記憶装置のコントローラが確認した後で、書き込みデータが不揮発性記憶装置に正常に書き込まれ

たことを示すために、ステータスメッセージを第１の記憶装置に送ることができる。書き込みデータが正常に書き込まれたことを示すステータスメッセージを受け取った後で、第１のビットマップを削除し、第２のビットマップを第１のビットマップとして指定することができる。

【０００８】

第２の記憶装置が利用可能でなかった期間の後で、第１のビットマップの内容を、回復ビットマップにコピーし、次にこれを用いて第１の記憶装置から第２の記憶装置にコピーされることになる第１の記憶装置の記憶領域を識別することができる。第１の記憶装置の識別された記憶領域を第２の記憶装置にコピーすることができ、新規に受け取られた書き込み要求を第３のビットマップ内に第２の記憶装置において蓄積することができる。

【０００９】

第２の記憶装置は、第１の記憶装置に対して上述したような１つ又はそれ以上の特徴及び機能を実行することができ、第１の記憶装置は第２の記憶装置に対して上述したような１つ又はそれ以上の特徴及び機能を実行することができる。

【００１０】

別の一般的な態様において、コンピュータシステムにおける第１の記憶装置のミラーコピーを第２の記憶装置に保持することは、関連するコントローラと、揮発性記憶装置と、不揮発性記憶装置とを含む第１の記憶装置において書き込み要求を受け取る段階と、第１の記憶装置で受け取られた書き込み要求を処理する段階と、関連するコントローラと揮発性記憶装置と不揮発性記憶装置とを含む第２の記憶装置において書き込み要求を受け取る段階と、第２の記憶装置において受け取られた書き込み要求を処理する段階とを含む。第２の記憶装置が書き込み要求を処理することができなくなる期間に入ろうとしていることを第２の記憶装置が判定した後で、第１の記憶装置のコントローラは、書き込み要求を指定する情報と共に第２の記憶装置に対して同期コミットメッセージを送り、第２の記憶装置のコントローラは、同期コミットメッセージを受け取った後で、指定された書き込み要求に関連するデータが第２の記憶装置の不揮発性記憶装置に書き込まれたことを確認する。同期コミットメッセージを送った後で、第１の記憶装置のコントローラは、ビットマップ内に新規の書き込み要求によって影響を受ける記憶の領域を蓄積する。第２の記憶装置が書き込み要求を再度処理できるようになると、第１の記憶装置から第２の記憶装置へコピーされることになる第１の記憶装置の記憶の記憶領域を識別するように、第１の記憶装置のコントローラはビットマップを使用して、第１の記憶装置の識別された領域の内容を第２の記憶装置にコピーする。

【００１１】

上述の手法の実施形態は、方法又はプロセス、装置又はシステム、或いはコンピュータがアクセス可能な媒体上のコンピュータソフトウエアを含むことができる。

【発明を実施するための最良の形態】

【００１２】

１つ又はそれ以上の実施形態の詳細が添付図面並びに以下の説明で記載される。他の特徴及び利点は、以下の説明、図面、及び請求項から明らかになろう。

各図面中、同一の参照記号は同一の要素を表す。

【００１３】

図１は、第１のデータ記憶装置１０５と第２のデータ記憶装置１１０を含むミラーセット１００のブロック図である。図１の実施形態において、データ記憶装置はディスクドライブである。他の実施形態では、データ記憶装置はディスクドライブのアレイ又は他の記憶装置とすることができる。

【００１４】

説明を簡単にするため、ディスクの１つはマスタディスクとされ、主データ記憶装置として機能し、他のディスクはスレーブディスクとされ、冗長バックアップとして機能する。双方のディスクがアクティブで且つ同じデータを含む場合には、マスタ／スレーブのステータスは２つのディスクに対して任意に割り当てることができる。実際に以下に述べる

同期化技法に対して、２－ディスクの実施形態では実際に２つのマスター－スレーブ関係が維持され、各ディスクは一方の関係ではマスタとして、他方の関係ではスレーブとして機能する。図１においては、ディスク１０５はマスタディスクとして指定され、ディスク１１０はスレーブディスクとして指定されている。

【００１５】

第１のＩ／Ｏ（「入力／出力」）コントローラ１１５は第１のディスク１０５と関連付けられ、第２のＩ／Ｏコントローラ１２０は第２のディスク１１０と関連付けられる。Ｉ／Ｏコントローラ１１５、１２０は、ディスク上のデータの読み込みと書き込みを制御する。

【００１６】

例えば、プロセッサとすることができるクライアント１２５は、同一の書き込み要求１３０を両方のＩ／Ｏコントローラに送る。各書き込み要求はデータを含む。加えて、連続参照番号などの基準ラベルは、各書き込み要求に関連付けられる。Ｉ／Ｏコントローラは、書き込み要求からのデータをそれぞれのディスクに書き込むので、通常の条件下では両方のディスクは同一のデータを含む。典型的には、各Ｉ／Ｏコントローラは書き込み要求を同じ順番で処理する。これを達成するために、Ｉ／Ｏコントローラは書き込み要求を基準ラベルの順番に処理するので、これはＩ／Ｏコントローラが同一順番の書き込み要求を受け取る必要がないことを意味する。

【００１７】

また、クライアント１２５は読み出し要求１３５をＩ／Ｏコントローラに送る。１つの実施形態において、両方のディスクが同一のデータを含む場合、マスタディスクだけが読み出し要求１３５に応答する。他の実施形態では、スレーブディスク又は両方のディスクが応答することができる。マスタディスクに障害が発生するか又はアクセス不能となった場合には、スレーブディスクがマスタディスクとして指定変更され、継続してデータをクライアント１２５に供給する。従って、ディスク１０５に障害が発生した場合にはディスク１１０がマスタディスクとなる。

【００１８】

ミラーセット１００のディスクは、ある時間期間の間に書き込み要求を処理することができない場合には、ピアのディスクとは不一致のデータを含むようになる。例えば、スレーブディスクがある時間期間の間に利用不能であった場合、スレーブディスクのデータはマスタディスクのデータとは異なるものとなる。ミラーセットのディスクが不一致になると、ミラーセットコピーを実行して、「正常な」データを有するディスクから不一致データを有するディスクにデータをコピーする。幾つかの大容量記憶装置においては、このプロセスは長時間を要する可能性があり、この間はミラーディスクが同一のデータを含まないことにより、システムのフォールト・トレランスのレベルが狭くなる。

【００１９】

システムのフォールト・トレランスレベルを改善するために、「良好な」データを有するディスクから、不一致データを有するディスクに保存されていなかったデータ部分だけをコピーすることによって、両方のディスクが同一データを含む状態（回復と呼ぶ場合がある）にミラーディスクを復元するのに必要な時間を短縮することができる。ディスクの部分だけをコピーするこのプロセスは、増分不一致コピー又は差分コピー（ここで、差分とは、一方のディスクに加えられており、他のディスクにはない変更を意味する）と呼ぶことができる。

【００２０】

一般に、増分不一致コピーは、１つ又はそれ以上のミラーディスクに対して行われるスレーブの変更によって行うことができ、その結果、利用不能期間の後で、「良好な」データだけを有するディスクから不一致データを有するミラーディスクに保存されていなかったデータをコピーすることによって、不一致データを有するミラーディスクを復元することができるようになる。一般に、ミラーディスクに対して加えられてきた変更の監視は、ミラーセット内の両方のミラーディスクが同一のデータを含むことが分かった時点（この

時点ではミラーディスクは同期しているとすることができる）の後で加えられた変更の推移を把握することが必要とされる。

【００２１】

同期化された時点の後にミラーディスクに加えられた変更の監視は、データが揮発性のディスクキャッシュには既に書き込まれているが、ミラーディスクの不揮発性記憶装置には未だ書き込まれていない時に、システム、サブシステム、又はプロセッサが書き込み要求を完了したものとして記録する際に問題となる可能性がある。２つ以上のＲＡＩＤ（Ｒｅｄｕｎｄｕｎｔ　Ａｒｒａｙ　ｏｆ　Ｉｎｅｘｐｅｎｓｉｖｅ　Ｄｉｓｋｓ）ディスクへの書き込みに要する時間が増大したことに起因して、書き込み要求がディスクキャッシュに出されてから、全てのデータの不揮発性ディスク記憶装置への書き込みを完了するまでの時間期間はかなりのものとなる可能性があるので、例えばＲＡＩＤ方式を使用して２つ以上のディスクにわたってミラーデータをストライピングした場合に、これは特に重大な問題となる恐れがある。

【００２２】

ディスクキャッシュをフラッシュしてキャッシュ内のデータをディスク記憶装置にコミットすることにより、特定の書き込み要求の基準ラベルまでミラーディスク上のデータを周期的に同期化することで（例えば各ディスクが同一のデータを含む場合）増分不一致コピーの効果を改善することができる。ディスクキャッシュをフラッシュすることにより、処理されていた全ての書き込み要求が不揮発性ディスク記憶装置内に確実に保存される。

【００２３】

ミラーセット１００内の各Ｉ／Ｏコントローラ１１５、１２０は、ビットマップ１５５、１５６内のディスクに対して加えられた変更を蓄積することによって、Ｉ／Ｏコントローラのそれぞれのディスクに対して行われた書き込み要求１３０の推移を把握する。ビットマップ１５５、１５６は、ディスクの各領域が書き込み要求１５０によって影響を受けたかどうかを表示するために１つ又はそれ以上のビットを使用するデータ構造である。この実施形態におけるビットマップはディスクに保存される。別の実施形態では、システムが停止するか、又はミラーセットに含まれていない不揮発性記憶装置にビットマップを保存するまで、揮発性メモリにビットマップを保存することができる。ビットマップによって得られる抽象性（又は粒度）のレベルは、ビットで表わされる記憶領域のサイズに基づく。一般に、各ビットは、単一の書き込み要求によって記憶装置の対応する領域に書き込まれるデータよりも実質的に大きなデータを表す。ここでビットマップ１５５、１５６はディスク変更ビットマップと呼ばれることがある。

【００２４】

ビットマップ及びディスクは、固有の識別子によって関連付けることができる。固有の識別子は、例えば、ビットマップとディスクが適用されるクライアント１２５のインスタンスを識別するディスク識別子を組み込むことができる。ビットマップとディスクを関連付けることにより、変更されたディスク領域を適切なディスクに確実にコピーすることができる。例えば、特定のビットマップと特定のディスク又はディスク上の特定のデータセットとの関連付けは、ミラーセットが取り外し可能ディスク（すなわち、コンピュータのハウジングユニットを開かずに取り外すことが可能なディスク）を含む場合に重要である。

【００２５】

マスタＩ／Ｏコントローラと呼ぶことができる１つのＩ／Ｏコントローラが、スレーブＩ／Ｏコントローラと呼ぶことができる他のＩ／Ｏコントローラに同期コミットメッセージ１６０を周期的に送る。図示し且つ以下に説明されるように、第１のＩ／Ｏコントローラ１１５がマスタＩ／Ｏコントローラであり、第２のＩ／Ｏコントローラ１２０がスレーブＩ／Ｏコントローラである。しかしながら、ディスク１１０がマスタであり、ディスク１０５がスレーブという関係で、第２のＩ／Ｏコントローラ１２０が同時にマスタＩ／Ｏコントローラとして作用する（第１のＩ／Ｏコントローラが同時にスレーブＩ／Ｏコントローラとして作用する）点に留意することは重要である。

【００２６】
同期コミットメッセージ１６０は、ミラーディスク上のデータが同期化されるようになる書き込み要求の基準ラベルを識別する。第１のＩ／Ｏコントローラ１１５は、ディスク変更ビットマップ１５５のバックアップコピー１６５を作成して同期化プロセスの間に障害が発生した場合に回復できるようにし、新しいディスク変更ビットマップを開始させて、次の同期化において使用されることになる後続の全ての書き込みを蓄積する。
【００２７】
第２のＩ／Ｏコントローラ１２０が、第１のＩ／Ｏコントローラ１１５によって送られた同期コミットメッセージを受け取ると、第２のＩ／Ｏコントローラは、該第２のＩ／Ｏコントローラが同期コミットメッセージにおいて識別された書き込み要求と全てのこれまでの書き込み要求とを既に処理したかどうかを判断する。処理していない場合、第２のＩ／Ｏコントローラは、同期化を開始する前に、当該書き込み要求と全てのこれまでの書き込み要求とが処理されるまで待機する。
【００２８】
第２のＩ／Ｏコントローラ１２０が、同期コミットメッセージで識別された書き込み要求と全てのこれまでの書き込み要求とを処理すると、又は同期コミットメッセージを受け取った時に、第２のＩ／Ｏコントローラが書き込み要求と全てのこれまでの書き込み要求とを既に処理していた場合には、第２のＩ／Ｏコントローラは、そのディスクコントローラキャッシュをフラッシュして処理された書き込み要求を不揮発性記憶装置にコミットする。キャッシュのフラッシュ化が正常である場合、第２のＩ／Ｏコントローラ１２０は第１のＩ／Ｏコントローラ１１５に確認メッセージ１７０を送る。フラッシュ及び同期化が正常であったことの確認を受け取るとすぐに、第１のＩ／Ｏコントローラ１１５は、このディスク変更ビットマップのバックアップコピー１６５をクリアする。同期化が正常ではない場合、又は所定の時間内に第１のＩ／Ｏコントローラ１１５が確認を受け取らない場合には、第１のＩ／Ｏコントローラ１１５は、ビットマップ１５５とバックアップ１６５とを（通常は論理和を取ることによって）結合し、結合されたビットマップを第２のディスク１１０の復元に使用する。
【００２９】
特定の時点からディスクに加えられた変更だけを蓄積する増分不一致コピープロセスは、ディスク障害が検出されると開始する（従って、利用不能の期間の間に加えられた変更だけを蓄積する）ことができ、又は、システムがアクティブである時は何時でも用いることができる（従って、システム動作中に常時ミラーセットに加えられた変更を蓄積する）。利用不能の期間だけ変更が蓄積される時には、利用不能の期間は、利用可能になった後でディスクを回復する際に有効となる増分不一致コピープロセスが利用不能になっている（これは、「正常な停止」と呼ぶことができる、プロセス全体でディスクが利用不能になる場合に行われることがある）ディスクのディスクコントローラのキャッシュフラッシュから開始する必要がある。
【００３０】
別の実施形態は、異なるディスク変化ビットマップの同期化時点後に加えられたディスク変化の蓄積を開始するのではなく、ディスク変化ビットマップにおいて特定の書き込み要求を削除することによる変化を蓄積することを含むことができる。これは回復時間を短縮することができる。
【００３１】
ミラーセット１００のＩ／Ｏコントローラ１１５と１２０のいずれかは、同期コミットプロセスを開始して、両方のディスクが特定の書き込み要求の基準ラベル全体を通じて同一のデータを含むよう保証することができる。利用不能期間の後で、最後の同期化以降変更されたディスク領域だけをコピーすることによって、同一データを保存するミラーコピーであるように不一致データを有するミラーディスクを復元することができる。
【００３２】
１つの実施形態では、日付（又は日付と時間）とミラーディスクの１つに対する特定の

データセットとの関連付けを含むことができる。これは、書き込み要求の基準ラベルが必ずしも固有のものでない場合に有益とすることができる。例えば、基準ラベルが、クライアントを制御するオペレーティングシステムのリセット時において、ある固定値（例えば１）から再起動する連続番号である場合、この書き込み要求の基準ラベルは固有ではない可能性がある。このような書き込み要求は、データセットを保存するディスクのディスクキャッシュがフラッシュされる時に、日付（又は日付と時間）をデータセットと関連付けることによって固有に識別することができる。別の方法として又は追加として、クライアントが再起動される時に、非固有の書き込み要求の基準ラベルの識別に役立つように、ミラーセットに新しいインスタンス番号を付与することができる。他の固有の識別子は、クライアント識別子、ミラーセット識別子、及びデータセット識別子を単独で又は組み合わせて含むことができる。

【００３３】

例証として図１では、ミラー装置として２つのディスクを使用してミラーデータのセットを保存しているが、増分不一致コピーの利点は、この特定の実施形態に限定されるものではなく、ＲＡＩＤ方式を含む別の数又は種類の記憶装置を伴う他の実施形態に対しても同様に適用可能である。例えば、別の実施形態は、３つ又はそれ以上のディスクをミラーリングすることができ、又はミラーディスクの複数のインスタンス化（例えば、４つのディスクを使用して同一のディスクに対して２つのミラーセットを提供することができる）を行うことができる。

【００３４】

図２を参照すると、プロセス２００は増分不一致追跡を使用して、ミラーディスクセットのディスクが、正常な停止を通じて利用不能になった期間中に、不一致になったミラーディスクセットへの同期化を復元する準備をする。図２で設定されたミラーディスクの実施形態は、別個のＩ／Ｏコントローラによって各々制御される２つのディスク記憶装置を有する。各Ｉ／Ｏコントローラは、プロセッサから同一の書き込み要求を受け取り、受け取った書き込み要求を起こった順番に処理する。基準ラベルは各書き込み要求と関連付けられ、書き込み要求の順序付けに使用される。別の実施形態では、特定の書き込み要求が処理されたか（完了したか）どうかを追跡することができる。これにより書き込み要求をバラバラの順序で（すなわち非順次で）処理することが可能となる。両方のディスクがアクティブである場合、各ディスクは同一のデータを含む。

【００３５】

プロセス２００は、ディスクの１つが利用不能の期間（ステップ２０５）に入っていると判定された時に開始される。この判定がなされると、利用不能になっているディスクのＩ／Ｏコントローラは、Ｉ／Ｏコントローラによって既に処理されている不揮発性記憶装置の書き込み要求に対してコミットするよう命令される（ステップ２１０）。利用不能となるディスクをスレーブディスク、アクティブなディスクをマスタディスクと呼び、これらに関連付けられたＩ／Ｏコントローラは、スレーブＩ／Ｏコントローラとマスタ Ｉ／Ｏコントローラと呼ぶことができる。マスタＩ／Ｏコントローラは、ディスク変更ビットマップ内のマスタディスクに加えられた変更の蓄積を開始し（ステップ２２０）、継続して書き込み要求を受け取って処理し（ステップ２２５）、ディスク変更ビットマップを更新し、処理された書き込み要求から結果として得られた全ての変更をマスタディスクに反映させる（ステップ２３０）。ディスク変更ビットマップの各ビットは、マスタディスクの領域を表す。他の実施形態では、ディスク変更ビットマップの各ビットによって表示されるディスクスペース量を変えることができる。

【００３６】

マスタＩ／Ｏコントローラは又、継続してスレーブディスクのステータスを監視する（ステップ２３５）。スレーブディスクが使用可能となり、新規の書き込み要求の処理を開始すると、マスタＩ／Ｏコントローラは、図３に関連して以下に説明するように、回復プロセス３００を開始する（ステップ２４０）。

【００３７】

図３を参照すると、回復プロセス３００は、ディスク変更ビットマップによって示されるように、マスタディスクのスレーブディスク部分へのコピーを含む。回復プロセスは、ミラーセットが継続して新規の書き込み要求を処理している間にアクティブなバックグラウンドプロセスとして行われる。回復プロセス３００は、マスタＩ／Ｏコントローラがディスク変更ビットマップのバックアップコピーを作成し、ディスク変更ビットマップのオリジナルバージョンを回復ビットマップとして指定することで開始する（ステップ３１０）。また、マスタＩ／Ｏコントローラは、新しいディスク変更ビットマップを開始して、後続の全ての変更をマスタディスクに蓄積する（ステップ３２０）。ディスク変更ビットマップと新しいディスク変更ビットマップのバックアップコピーは、回復プロセスの間に障害が発生した場合の回復を可能にする。

【００３８】

マスタＩ／Ｏコントローラは、回復ビットマップ内の各ビットを検査し（ステップ３３０）、対応するマスタディスクの領域が既に変更されたことをビットが示すかどうかを判定する（ステップ３４０）。示さない場合、マスタＩ／Ｏコントローラは回復ビットマップの次のビットの検査に進む（ステップ３４５）。マスタディスク領域が既に変更されたことをビットが示す場合、マスタＩ／Ｏコントローラは、後続の書き込み要求が、スレーブディスクの対応するディスク領域を変更したかどうかを判定する（ステップ３４５）。

【００３９】

後続の書き込み要求が、対応するスレーブディスク領域を既に変更していた場合には、マスタＩ／Ｏコントローラは、後続の書き込み要求によって変更されていなかったスレーブディスク領域の部分に相当する、マスタディスク領域の部分だけをコピーする（ステップ３５０）。マスタＩ／Ｏコントローラは、スレーブＩ／Ｏコントローラに、回復プロセスの間にスレーブディスクにより処理される書き込み要求のリストを保持させることにより、変更されていない領域を識別することができ、リストの各エントリは修正された実メモリ部分を識別する。或いは、スレーブＩ／Ｏコントローラは、より微細な粒度を有するディスク変更ビットマップを保持することができるので、ビットマップの各々のビットは、書き込み要求の修正が許可されているディスクの最小の部分に対応するようにする。スペースを保護するために、スレーブＩ／Ｏコントローラは、粒度が変化するビットマップを保持して、ディスクの修正された部分に対してだけ微細な粒度のマップが保持されるようにすることができる。

【００４０】

スレーブディスク領域に対して後続の変更が何も加えられなかった場合、マスタＩ／Ｏコントローラは、マスタディスクの全領域をスレーブディスクにコピーする（ステップ３５５）。

【００４１】

マスタＩ／Ｏコントローラは、コピーされているデータの部分を修正し、後続の書き込み要求によって上書きされることになるデータ書き込みの潜在的な非能率さを排除する（ステップ３４５－３５５）。例えば、書き込み要求ＷＲ－１０２がディスク領域１２に保存されたデータの部分を変更し、ＷＲ－１５５もまた、ディスク領域１２の別の部分に保存されたデータを変更する場合には、ディスク領域１２へのデータを書き込むプロセスは、各書き込み要求に必要とされる領域１２の部分だけを変更することができる。

【００４２】

追加又は代替として、スレーブＩ／Ｏコントローラは、コピーされているデータの部分を修正することができる。例えば、スレーブＩ／Ｏコントローラが、マスタディスクからコピーされたデータによって更新されることになる同一のディスク領域を修正する新規の書き込み要求を受け取った場合には、スレーブＩ／Ｏコントローラは、マスタディスクからコピーされているデータの部分を修正することができる。

【００４３】

マスタディスク領域（又はその部分）をスレーブディスクにコピーした後、マスタＩ／Ｏコントローラは、回復ビットマップ内の別のビットを検査する必要があるかどうかを判

断し（ステップ３６０）、必要な場合には次のビットを検査する（ステップ３３０）。

【００４４】

マスタＩ／Ｏコントローラが、回復ビットマップ内の全てのビットが検査されたことを判定した時に回復が完了する。終了すると、コピーされたデータをスレーブディスクにコミットするために、マスタＩ／Ｏコントローラは、任意選択的にスレーブディスク・キャッシュをフラッシュする同期化を開始することができる（ステップ３７０）。マスタＩ／Ｏコントローラが、後続の同期化又はフラッシュ化が正常でないと判定する（ステップ３７５）と、マスタＩ／Ｏコントローラはディスク変更ビットマップのバックアップコピーを新しいディスク変更ビットマップと結合し（通常は論理和を取ることによって）（ステップ３８０）結合されたディスク変更ビットマップを使用して回復プロセス３００を繰り返す。スレーブＩ／Ｏコントローラの同期化とスレーブディスクコントローラのフラッシュ化が正常である場合には、マスタＩ／Ｏコントローラはバックアップディスク変更ビットマップをクリアする（ステップ３９０）。

【００４５】

図２に関連して説明された実施形態は、マスタディスクからスレーブディスクにディスク領域をコピーする時に粒度のレベルを変更するが、別の実施形態では、後続の書き込み要求によって領域の部分が上書きされるかどうかに関係なく、常に変更された領域全体をコピーすることができる。図３の実施形態は、回復ビットマップを回復の間に２度使用できないように処理する。回復プロセスの間に障害からの回復を可能とするために、ビットを処理する前にディスク変更ビットマップのバックアップコピーを行う（ステップ３１０）。別の実施形態では、回復ビットマップは回復の間に破棄されることはなく、回復ビットマップ自体を使用することによって、回復プロセス障害から回復することができる。この実施形態は、ビットを処理する前にディスク変更ビットマップのコピーを行うことはない（ステップ３１０）。別の代替実施形態は、スレーブディスクが使用可能に戻った後で、且つ回復プロセスが正常に完了する前に行われたマスタディスク変更を蓄積するために新しいディスク変更ビットマップを使用することはない。

【００４６】

図４－図６を参照すると、増分不一致コピープロセスは、ミラーセットを使用する時は常にアクティブとすることができる。図４－図６のミラーセットの実施形態は、図２に関連して説明された方法で書き込み要求を受け取る２つのディスク記憶装置と２つのＩ／Ｏコントローラとを有する。

【００４７】

各Ｉ／Ｏコントローラは、ディスク変更ビットマップ内のディスクに加えられた変更を蓄積することによって、Ｉ／Ｏコントローラのディスクに対して加えられた書き込み要求の推移を把握する。１つのＩ／Ｏコントローラ（マスタＩ／Ｏコントローラと呼ばれる）は、周期的に同期コミットメッセージを別のＩ／Ｏコントローラ（スレーブＩ／Ｏコントローラと呼ばれる）に送り、周期的な同期化プロセスを開始する。図４はマスタＩ／Ｏコントローラによって実行される周期的な同期化プロセスを示す。図５は、スレーブＩ／Ｏコントローラによって実行される周期的な同期化プロセスを示す。図６は、同一データを有するミラーコピーとなる、不一致データを有するミラーディスクを復元するためのプロセスを示す。

【００４８】

図４を参照すると、マスタＩ／Ｏコントローラは、スレーブＩ／Ｏコントローラと周期的な同期化を行うためのプロセス４００を開始する。プロセス４００は、マスタＩ／Ｏコントローラが、プロセッサから書き込み要求を受け取って処理し（ステップ４１０）、ディスク変更ビットマップ内のマスタディスクに加えられた変更を蓄積する（ステップ４１５）ことで開始する。マスタＩ／Ｏコントローラは、ミラーセットを同期化すべきかどうかを判断する（ステップ４２０）。マスタＩ／Ｏコントローラは、例えば、最後の同期化から所定時間が経過した後、最後の同期化から所定の数の書き込み要求が処理された後、又は２つのミラーディスク間の増分不一致が一定割合となった後で、同期コミットメッセ

ージを作成することができる。同期化が要求される時を判定する際に、ミラーディスク間で同期化されていないデータ量に対して同期化周波数（これは、不揮発性記憶装置にディスクキャッシュが書き込まれている時間の間はディスクキャッシュのフラッシュ化が全ての書き込み要求の処理を停止するので、システムの性能を低下させる可能性がある）を平衡させることができる（これは、ミラーセットに対して同一のデータを復元するための増分不一致コピーを行うのにより長い時間を要する可能性がある）。

【００４９】

ミラーセットを同期化すべきであることをマスタＩ／Ｏコントローラが判断すると、マスタＩ／Ｏコントローラは、同期コミットメッセージをスレーブＩ／Ｏコントローラに送る（ステップ４３０）。同期コミットメッセージは、ミラーディスク上のデータが同期化されることになる書き込み要求の基準ラベルを識別する。マスタＩ／Ｏコントローラは、ディスク変更ビットマップのバックアップコピーを作成（ステップ４３５）し、同期化プロセスで障害が発生した場合に回復できるようにし、新しいディスク変更ビットマップを開始し（ステップ４４０）て、次の同期化において使用するためにこの時点からマスタＩ／Ｏコントローラのディスクに加えられるディスク変更を蓄積する。マスタＩ／Ｏコントローラは、継続してプロセッサから書き込み要求を受け取って処理を行い（ステップ４４５）、マスタディスクの全ての変更を反映させるよう新しいディスク変更ビットマップを更新する（ステップ４５０）。

【００５０】

スレーブＩ／Ｏコントローラによるキャッシュフラッシュと同期化とが正常であった（ステップ４５５）ことの確認を受け取ると、マスタＩ／Ｏコントローラは、バックアップディスク変更ビットマップをクリアし（ステップ４６０）、同期化のプロセスが終了する。

【００５１】

或いは、例えば、マスタＩ／Ｏコントローラが所定の時間内にスレーブＩ／Ｏコントローラから確認メッセージを受け取らなかったか、又は同期化に障害が発生したというメッセージを受け取ったことにより、マスタＩ／Ｏコントローラは同期化が正常ではなかったことを判断することができる（ステップ４５５）。この場合には、マスタＩ／Ｏコントローラは、バックアップディスク変更ビットマップと新しいディスク変更ビットマップとを結合し（通常、論理和を取ることによって）（ステップ４７０）、スレーブＩ／Ｏコントローラとこれに関連するディスクとが動作可能であることを判断すると、結合されたディスク変更ビットマップを使用して、図３に関連して説明されたような回復プロセス３００を開始し、マスタディスクからスレーブディスクにどのディスク領域をコピーすべきかを導く（ステップ４７５）。

【００５２】

図５を参照すると、プロセス５００は、ミラーディスク上のデータが同期化されることになる書き込み要求基準ラベルを識別する同期コミットメッセージを、スレーブＩ／Ｏコントローラが受け取った時に開始する（ステップ５１０）。スレーブＩ／Ｏコントローラは、同期コミットメッセージと全てのこれまでの書き込み要求とにおいて識別された書き込み要求を既に処理したかどうかを判定する（ステップ５２０）。処理していない場合には、スレーブＩ／Ｏコントローラは、同期化を開始する前に、該書き込み要求と全てのこれまでの書き込み要求が処理されるまで待機する。

【００５３】

スレーブＩ／Ｏコントローラが、同期コミットメッセージで識別された書き込み要求と、全てのこれまでの書き込み要求とを処理すると、又は同期コミットメッセージが受け取られた時にスレーブＩ／Ｏコントローラが該書き込み要求と全てのこれまでの書き込み要求とを既に処理している場合には、スレーブディスクコントローラは、そのキャッシュをフラッシュして、処理された書き込み要求を不揮発性のディスク記憶装置にコミットし（ステップ５３０）、キャッシュフラッシュが正常であったかどうかを判定する（ステップ５４０）。キャッシュフラッシュが正常であった場合、スレーブＩ／Ｏコントローラは、

マスタＩ／Ｏコントローラに確認メッセージを送る（ステップ５５０）。キャッシュフラッシュが正常でなかった場合には、スレーブＩ／Ｏコントローラは、マスタＩ／Ｏコントローラに障害発生のメッセージを送る（ステップ５６０）。マスタＩ／Ｏコントローラに適切なメッセージを送った後、スレーブＩ／Ｏコントローラはプロセス５００を終了する。

【００５４】

図６は、同一データを保存するミラーコピーであるように不一致データを有するミラーディスクを復元するためのプロセス６００を示す。以下の説明は、ミラーセット内のディスクの１つ（スレーブディスク）がこれまでに障害が発生したか、或いは利用不能となっており、且つミラーセットが同期化された前回以降、他のアクティブなディスク（マスタディスク）に加えられた全ての変更を含むディスク変更ビットマップが存在することを仮定している。これは、例えば図４－図５に関連して説明されたプロセスを実行することによって達成することができる。

【００５５】

スレーブディスクが利用可能でない場合、マスタＩ／Ｏコントローラは、継続してプロセッサから書き込み要求を受け取って処理し（ステップ６１０）、最後の同期化以降に加えられたディスク変更を追跡するディスク変更ビットマップのマスタディスクに対して加えられた変更を蓄積する（ステップ６２０）。マスタＩ／Ｏコントローラが、スレーブディスクが回復していて書き込み要求処理の開始が可能であると判断する（ステップ６３０）と、マスタＩ／Ｏコントローラは、ディスク変更ビットマップを使用して、図３に関連して説明されたような回復プロセス３００を開始し、マスタディスクのデータと同一のデータを含むようにスレーブディスクを復元する（ステップ６４０）。

【００５６】

プロセス４００、５００及び６００を実行することによって達成された増分不一致コピーは、プロセス２００を実行することによって達成されたものとは異なる。特にプロセス４００－６００は、ディスク変更ビットマップがミラーセットのアクティブな間に更新されるので、ディスクの１つに予期しないディスク障害又はコントローラ障害が発生している間にミラーディスクセットを再確立するのに有効である。プロセス２００は、次のディスク利用不能期間の警告が、ディスクキャッシュのフラッシュ実行を可能にするのに十分であり、ディスク変更ビットマップに蓄積されるようになる残りのアクティブなディスクに対する変更が開始される際、ミラーディスクセットの再確立だけに有効である。しかしながら、プロセス２００は特定の時間に実行されるだけであるので、オーバーヘッドの処理はプロセス４００－６００よりも大幅に少ない結果となる。

【００５７】

図７を参照すると、回復プロセス７００は、周期的な同期化の実行と、ディスク変更ビットマップによって示されるようなマスタディスクの部分のスレーブディスクへのコピーとを含む。回復プロセス７００は、マスタＩ／Ｏコントローラがディスク変更ビットマップのバックアップコピーを作成して、ディスク変更ビットマップのオリジナルのバージョンを回復ビットマップとして指定（ステップ７１０）して開始する。マスタＩ／Ｏコントローラはまた、新しいディスク変更ビットマップを開始して、後続の全ての変更をマスタディスクに蓄積する（ステップ７２０）。

【００５８】

図３に関連して説明されたように、マスタＩ／Ｏコントローラは、回復ビットマップの各ビットを検査して（ステップ７３０）、マスタディスク領域が変更されたことをビットが示す場合、マスタディスクは、該マスタディスクの変更された部分をスレーブディスクにコピーする（ステップ７３５）。

【００５９】

図４に関連して上述したように、マスタＩ／ＯコントローラはスレーブＩ／Ｏコントローラとの同期化プロセスを周期的に開始する。特に、ミラーセットを同期化すべきであることをマスタＩ／Ｏコントローラが判断する（ステップ７４０）と、マスタＩ／Ｏコント

ローラは、同期コミットメッセージをスレーブＩ／Ｏコントローラに送り、ディスク変更ビットマップのコピーを作成し、新しいディスク変更ビットマップを開始して、この時点からマスタＩ／Ｏコントローラのディスクに加えられる変更を蓄積する（ステップ７４５）。

【００６０】

スレーブディスクコントローラによるキャッシュフラッシュと同期化とが正常であった（ステップ７５０）という確認を受け取ると、マスタＩ／Ｏコントローラは、スレーブディスクに正常にコピーされていたマスタディスクの領域を示すビットを、回復ビットマップから除外する（ステップ７５５）。マスタＩ／Ｏコントローラは、例えば回復プロセスの間にマスタディスクによって処理されたビットのリストを保持し、リストされたビットを回復ビットマップから削除することによって、これを達成することができる。しかしながら、マスタＩ／Ｏコントローラが、同期化が正常でなかったと判定する（ステップ７５０）と、マスタＩ／Ｏコントローラは、バックアップディスク変更ビットマップと新しいディスク変更ビットマップとを結合し（ステップ７６０）、結合されたディスク変更ビットマップを使用して、図３に関連して説明された回復プロセス３００を開始する（ステップ７６５）。

【００６１】

マスタＩ／Ｏコントローラが、回復ビットマップの全てのビットを検査したと判定する（ステップ７７０）と、回復プロセスを終了し、バックアップディスク変更のビットマップをクリアする（ステップ７７５）。実施形態は、方法又はプロセス、装置又はシステム、若しくはコンピュータ媒体上のコンピュータソフトウエアを含むことができる。添付の請求項の精神及び範囲から逸脱することなく、種々の変更が可能であることは理解されるであろう。例えば、開示された手法のステップを別の順序で実行する場合、及び／又は開示されたシステムの構成要素を別の方法で結合する場合、及び／又は別の構成要素によって置き換えられるか補足される場合には、有利な結果が得られるであろう。

【図面の簡単な説明】

【００６２】

【図１】ミラードライブシステムのブロック図である。

【図２】ミラーディスク間の差異を監視するプロセスのフローチャートである。

【図３】不一致になったミラーディスクのセットに対する同期性を回復するためのプロセスのフローチャートである。

【図４】マスタ入出力コントローラによって実行される周期的な同期化プロセスを示すフローチャートである。

【図５】スレーブ入出力コントローラによって実行される周期的な同期化プロセスを示すフローチャートである。

【図６】同一のデータを有するミラーコピーであるように、不一致データを有するミラーディスクを復元するためのプロセスのフローチャートである。

【図７】回復プロセスの間に実行される周期的な同期化を示すフローチャートである。

【符号の説明】

【００６３】

１１５　Ｉ／Ｏコントローラ１
１２０　Ｉ／Ｏコントローラ２
１２５　クライアント
１５５　ビットマップ
１５６　ビットマップ
１６０　同期コミットメッセージ
１６５　バックアップビットマップ
１７０　確認メッセージ

【図１】

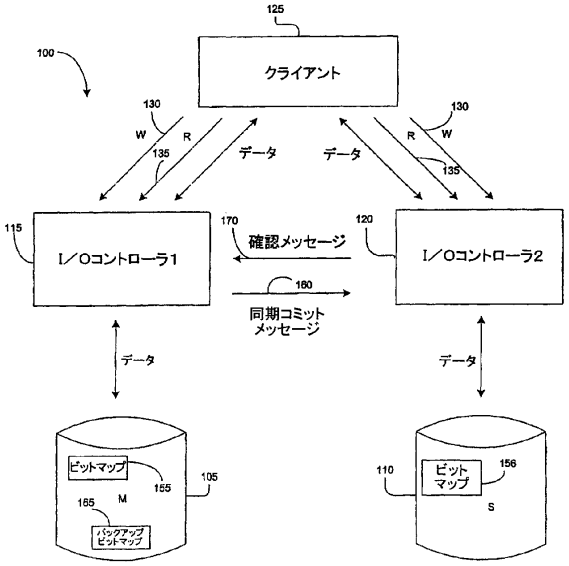

FIG. 1

【図２】

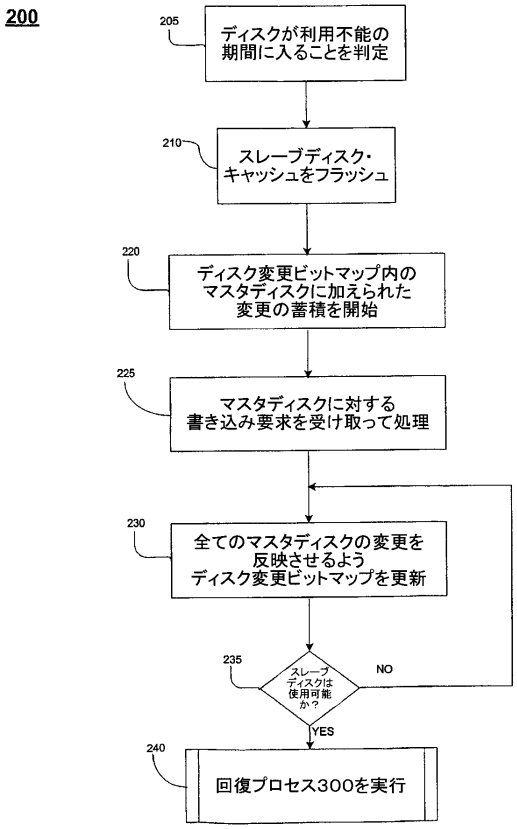

FIG. 2

【図３】

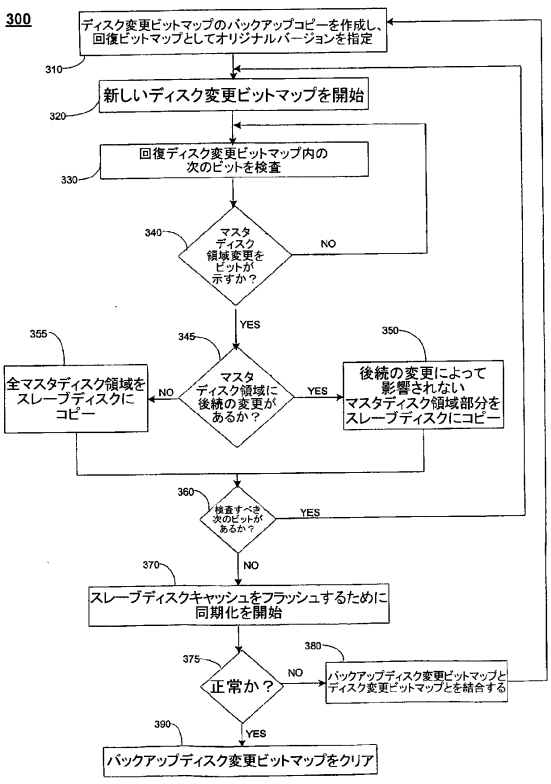

FIG. 3

【図４】

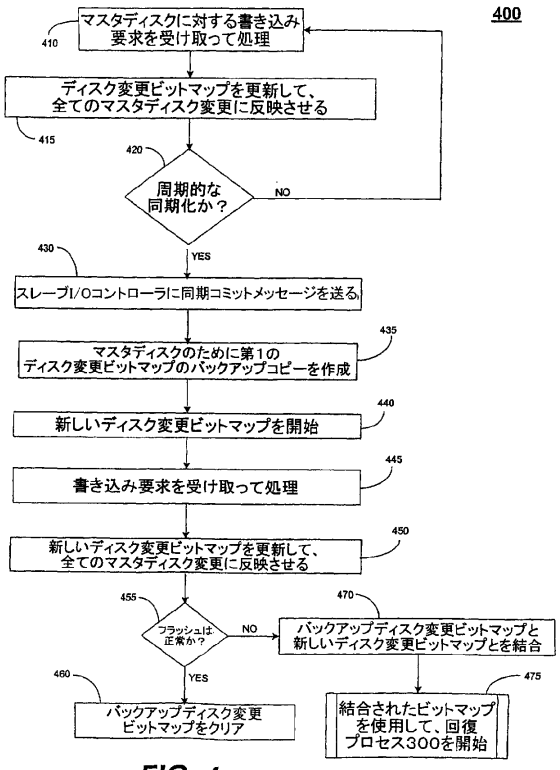

FIG. 4

【図５】

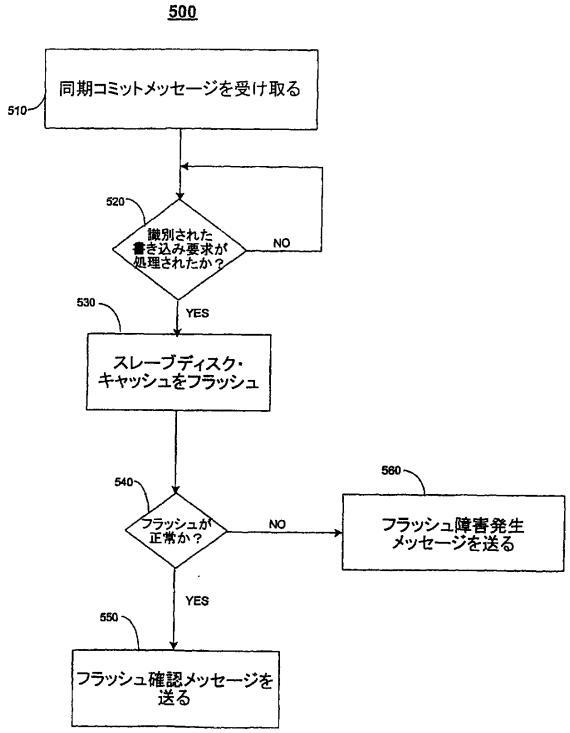

FIG. 5

【図６】

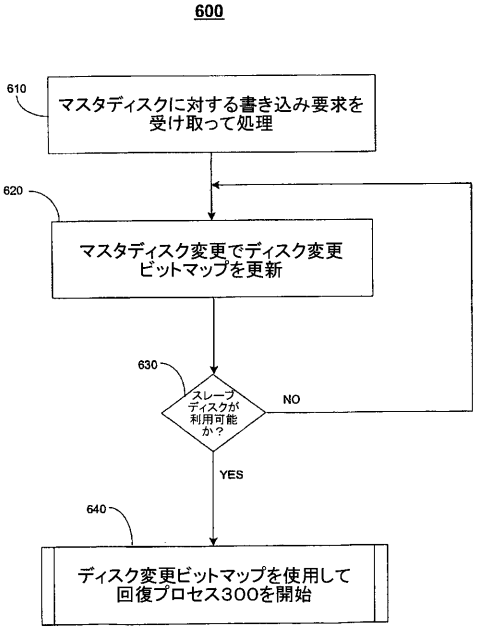

FIG. 6

【図７】

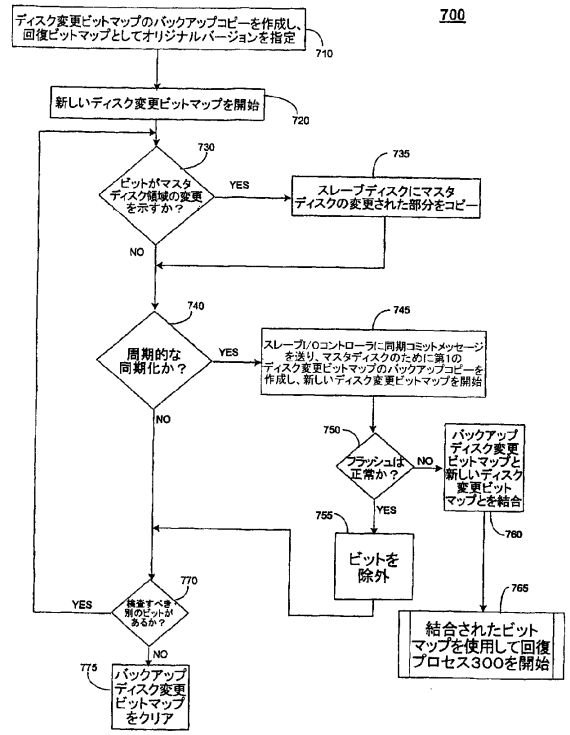

FIG. 7

フロントページの続き

(72)発明者 トレンブレイ グレン エイ
アメリカ合衆国 マサチューセッツ州 ０１５６８ アップトン サウス ストリート １３９
(72)発明者 リヴィール ポール エイ
アメリカ合衆国 マサチューセッツ州 ０１５１９ グラフトン ストラットン ロード １２
(72)発明者 カマン チャールズ エイチ
アメリカ合衆国 マサチューセッツ州 ０１７７３ リンカーン オーク メドー ロード １０
(72)発明者 グラナン ゲアリー
アメリカ合衆国 マサチューセッツ州 ０１７１９ ボックスボロー レオナード ロード ３５

審査官 上嶋 裕樹

(56)参考文献 特表２００１－５０１００２（ＪＰ，Ａ）
特開平２－１６６５０９（ＪＰ，Ａ）
国際公開第０１／０４０９５２（WO，Ａ１）

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