(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)特許出願公表番号
特表2005-526298
(P2005-526298A)
(43)公表日　平成17年9月2日(2005.9.2)

(51)Int.Cl.[7]	F I			テーマコード（参考）
G06F 12/00	G06F	12/00	533J	5B082
	G06F	12/00	514M	
	G06F	12/00	517	

審査請求 未請求　予備審査請求 有　（全 31 頁）

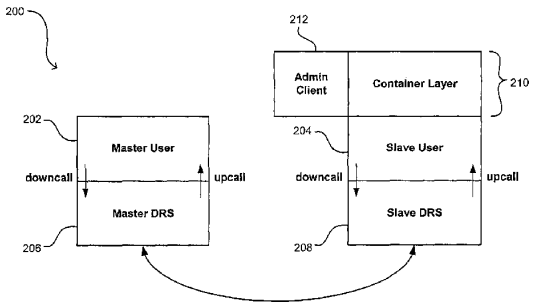
(21)出願番号	特願2003-514370 (P2003-514370)
(86)(22)出願日	平成14年7月15日 (2002.7.15)
(85)翻訳文提出日	平成16年3月15日 (2004.3.15)
(86)国際出願番号	PCT/US2002/022366
(87)国際公開番号	WO2003/009092
(87)国際公開日	平成15年1月30日 (2003.1.30)
(31)優先権主張番号	60/305,986
(32)優先日	平成13年7月16日 (2001.7.16)
(33)優先権主張国	米国(US)
(31)優先権主張番号	607305,978
(32)優先日	平成13年7月16日 (2001.7.16)
(33)優先権主張国	米国(US)
(31)優先権主張番号	09/975,590
(32)優先日	平成13年10月11日 (2001.10.11)
(33)優先権主張国	米国(US)

(71)出願人　500105160
ビーイーエイ　システムズ，　インコーポレイテッド
ＢＥＡ　Ｓｙｓｔｅｍｓ，　Ｉｎｃ．
アメリカ合衆国　カリフォルニア　９５１３１，　サン　ノゼ，　ノース　ファースト　ストリート　２３１５
２３１５　Ｎｏｒｔｈ　Ｆｉｒｓｔ　Ｓｔｒｅｅｔ，　Ｓａｎ　Ｊｏｓｅ，　ＣＡＬＩＦＯＲＮＩＡ　９５１３１　Ｕ．Ｓ．Ａ．
(74)代理人　100082005
弁理士　熊倉　禎男
(74)代理人　100067013
弁理士　大塚　文昭

最終頁に続く

(54)【発明の名称】データ複製プロトコル

(57)【要約】
一フェーズ法又は二フェーズ法を用いて、ネットワークを通じてデータを複製することができる。一フェーズ法の場合は、各々のスレーブがマスターからデルタを要求できるように、データのオリジナル・コピーを含むマスター・サーバが、該データの現在の状態についてのバージョン番号をネットワークで送信する。要求されたデルタは、スレーブを適切なバージョンのデータに更新するために必要なデータを含む。二フェーズ法の場合は、マスター・サーバは、情報のパケットを各々のスレーブに送信する。各々のスレーブがコミットを処理できる場合には、スレーブにより情報のパケットをコミットすることができる。



【特許請求の範囲】
【請求項１】
ネットワークを通じてマスター・サーバからスレーブ・サーバにデータを複製する方法であって、
前記マスター・サーバ上に格納された前記データの変更に関し、該データの現在の状態についてのバージョン番号を含む情報のパケットを、該マスター・サーバから前記スレーブ・サーバに送信し、
前記スレーブ・サーバ上の前記データが前記パケット内に含まれる前記バージョン番号に対応するように更新されたかどうかを、該スレーブ・サーバが判断することを可能にし、
前記スレーブ・サーバ上の前記データが前記パケット内に含まれる前記バージョン番号に対応しない場合には、該スレーブ・サーバを更新するのに必要とされる情報を含むデルタが、前記マスター・サーバから該スレーブ・サーバに送信されることを要求する、
段階を含むことを特徴とする方法。
【請求項２】
前記データのオリジナル・コピーを前記マスター・サーバ上に格納する段階をさらに含むことを特徴とする請求項１に記載の方法。
【請求項３】
各々のスレーブ・サーバについて、ローカル・ディスク上の前記データを永続的にキャッシングする段階をさらに含むことを特徴とする請求項１に記載の方法。
【請求項４】
前記データが変更された場合には、前記マスター・サーバ上の前記データの現在の状態についての固有のバージョン番号を判断する段階をさらに含むことを特徴とする請求項１に記載の方法。
【請求項５】
ネットワークを通じてマスター・サーバからスレーブ・サーバにデータを複製する方法であって、
前記マスター・サーバ上に格納された前記データの現在の状態に関するバージョン番号を、該マスター・サーバから前記スレーブ・サーバに送信し、
前記スレーブ・サーバが前記マスター・サーバから送信された前記バージョン番号に対応する前記データの現在の状態を反映するように更新されたかどうかを、該スレーブ・サーバが判断することを可能にし、
前記スレーブ・サーバが前記マスターにより送信された前記バージョン番号に対応しない場合には、該スレーブ・サーバを更新するのに必要とされる情報を含むデルタが、該マスター・サーバから該スレーブ・サーバに送信されることを要求する、
段階を含むことを特徴とする方法。
【請求項６】
前記マスター・サーバから前記スレーブ・サーバに前記デルタを送信する段階をさらに含むことを特徴とする請求項５に記載の方法。
【請求項７】
前記スレーブ・サーバへの前記デルタをコミットする段階をさらに含むことを特徴とする請求項５に記載の方法。
【請求項８】
前記デルタをコミットした後、前記スレーブ・サーバの前記バージョン番号を更新する段階をさらに含むことを特徴とする請求項５に記載の方法。
【請求項９】
前記バージョン番号を前記マスター・サーバからスレーブ・サーバに定期的に送信する段階をさらに含むことを特徴とする請求項５に記載の方法。
【請求項１０】
前記スレーブ・サーバが前記バージョン番号の受信を確認するまで、前記バージョン番

号をスレーブ・サーバに送信する段階をさらに含むことを特徴とする請求項５に記載の方法。
【請求項１１】
スレーブ・サーバを更新するのに必要な前記バージョン番号を有するデータを含ませる段階をさらに含むことを特徴とする請求項５に記載の方法。
【請求項１２】
データが受け取るとすぐに、前記スレーブを更新するのに必要な該データをコミットする段階をさらに含むことを特徴とする請求項１１に記載の方法。
【請求項１３】
前記マスター・サーバから前記デルタを送信する前に、該デルタの範囲を判断する段階をさらに含むことを特徴とする請求項５に記載の方法。
【請求項１４】
マスター・サーバ及び少なくとも１つのスレーブ・サーバを含むネットワークを通じてデータを複製する方法であって、
前記マスター・サーバ上に格納された前記データの変更に関し、該データの現在の状態についてのバージョン番号を含み、さらに、該データの前の変更及び前の変更の各々についてのバージョン番号に関する情報のパケットを、ネットワークで該マスター・サーバからスレーブ・サーバの各々に送信し、
前記スレーブ・サーバが前記バージョン番号に対応するように更新されたかどうかを、各々のスレーブ・サーバが判断することを可能にし、
前記スレーブ・サーバが前の変更を逃さなかった場合には、各々のスレーブ・サーバが前記情報をコミットすることを可能にし、
前記スレーブ・サーバが前記情報のパケットをコミットする前に、前の変更が、前記マスター・サーバから該スレーブ・サーバに送信されることを、前の変更を逃した各々のスレーブ・サーバが要求することを可能にする、
段階を含むことを特徴とする方法。
【請求項１５】
スレーブ・サーバへの前記情報のパケットをコミットする段階をさらに含むことを特徴とする請求項１４に記載の方法。
【請求項１６】
スレーブ・サーバが前記更新をコミットできない場合には、前記情報のパケットの前記コミットを中止する段階をさらに含むことを特徴とする請求項１４に記載の方法。
【請求項１７】
前記マスター・サーバから前記デルタを送信する前に、該デルタの範囲を判断する段階をさらに含むことを特徴とする請求項１４に記載の方法。
【請求項１８】
前記デルタの前の変更の各々の範囲を含ませる段階をさらに含むことを特徴とする請求項１４に記載の方法。
【請求項１９】
マスター・サーバ及び少なくとも１つのスレーブ・サーバを含むネットワークを通じてデータを複製する方法であって、
マスター・サーバ上に格納された前記データの変更に関し、前の状態についての前のバージョン番号及び該データの新しい状態についての新しいバージョン番号を含み、さらに、該データの前の変更及び各々の前の変更についての前のバージョン番号に関連する情報のパケットを、ネットワークで前記マスター・サーバから各々のスレーブ・サーバに送信し、
前記スレーブ・サーバ上の前記データが前記パケット内に含まれる前記前のバージョン番号に対応するかどうかを、各々のスレーブ・サーバが判断することを可能にし、
前記スレーブ・サーバ上の前記データが前記パケット内に含まれる前記前のバージョン番号に対応する場合には、各々のスレーブ・サーバが前記情報のパケットをコミットする

ことを可能にする、
段階を含み、前記コミットが、前記スレーブ・サーバの前記バージョンを前記新しいバージョン番号にも更新し、
前記スレーブ・サーバが前記情報のパケットをコミットする前に、前記スレーブを前記前のバージョン番号に更新するのに必要な前記情報を含むデルタが、前記マスター・サーバから送信されることを、該前のバージョン番号に対応しないスレーブ・サーバの各々が要求することを可能にする、
段階が設けられたことを特徴とする方法。
【請求項２０】
マスター・サーバ及び少なくとも１つのスレーブ・サーバを含むネットワークを通じてデータを複製する方法であって、
マスター・サーバ上に格納された前記データの変更に関し、前の状態についてのバージョン番号及び該データの新しい状態についてのバージョン番号を含み、さらに、該データの前の変更及び前の変更の各々についてのバージョン番号に関する情報のパケットを、ネットワークで前記マスター・サーバからスレーブ・サーバの各々に送信し、
前記スレーブ・サーバ上の前記データが前記パケット内に含まれる前記前のバージョン番号に対応するかどうかを、各々のスレーブ・サーバが判断することを可能にし、
前記スレーブ・サーバ上の前記データが前記パケットに含まれる前記前のバージョン番号に対応する場合には、各々のスレーブ・サーバが前記情報のパケットをコミットすることを可能にする、
段階を含み、前記コミットが、前記スレーブ・サーバの前記バージョンを前記新しいバージョン番号にも更新し、
前記スレーブを前記新しいバージョン番号に更新するのに必要な前記情報を含むデルタが、前記マスター・サーバから送信されることを、前記前のバージョン番号に対応しないスレーブ・サーバの各々が要求することを可能にする、
段階が設けられたことを特徴とする方法。
【請求項２１】
ネットワークを通じて、マスター・サーバかた、少なくとも１つのスレーブ・サーバにデータを複製する方法であって、
前記マスター・サーバ上に格納された前記データの変更に関し、前記データの現在の状態についてのバージョン番号を含む情報のパケットを、前記マスター・サーバからスレーブ・サーバに送信し、
スレーブ・サーバへの前記情報のパケットを受信し、
前記スレーブ・サーバが前記パケット内に含まれる前記バージョン番号に対応するように更新されたかどうか、さらに、該パケット内に含まれる該バージョン番号に対応するように更新する必要がある場合には、該スレーブ・サーバが該情報のパケットを処理することができるかどうかを、該スレーブ・サーバが判断することを可能にし、
前記スレーブ・サーバが更新されることを必要とするかどうか、及び該スレーブ・サーバが前記更新を処理することができるかどうかを示す信号を、該スレーブ・サーバから前記マスター・サーバに送信し、
前記スレーブ・サーバが前記パケット内に含まれる前記情報にコミットすべきであるかどうかを示す応答信号を、前記マスター・サーバから該スレーブ・サーバに送信し、
前記応答信号により前記スレーブ・サーバが前記パケット内に含まれる前記情報にコミットすべきであることが示された場合には、前記スレーブ・サーバへの該情報のパケットをコミットする、
段階を含むことを特徴とする方法。
【請求項２２】
少なくとも１つのスレーブ・サーバの各々が前記データをコミットできるかどうかを判断する段階をさらに含むことを特徴とする請求項２１に記載の方法。
【請求項２３】

少なくとも１つのスレーブ・サーバの各々が、応答を前記マスター・サーバに送り返したかどうかを判断する段階をさらに含むことを特徴とする請求項２１に記載の方法。
【請求項２４】
少なくとも１つのスレーブ・サーバのいずれかが前記データをコミットできるかどうかを判断する段階をさらに含むことを特徴とする請求項２１に記載の方法。
【請求項２５】
少なくとも１つのスレーブ・サーバの各々が前記コミットを処理できる場合に限り、前記データをコミットする段階をさらに含むことを特徴とする請求項２１に記載の方法。
【請求項２６】
少なくとも１つのスレーブ・サーバの各々が前記コミットを処理できない場合に限り、前記データを中止する段階をさらに含むことを特徴とする請求項２１に記載の方法。
【請求項２７】
前記コミットを処理することができるそれらのスレーブへの前記データをコミットする段階をさらに含むことを特徴とする請求項２１に記載の方法。
【請求項２８】
前記コミットを処理することができなかった前記少なくとも１つのスレーブ・サーバのいずれかに前記更新をマルチキャストする段階をさらに含むことを特徴とする請求項２１に記載の方法。
【請求項２９】
前記新しいバージョン番号を、前記コミットを処理することができなかった前記少なくとも１つのスレーブ・サーバのいずれかに一定間隔で送信する段階をさらに含むことを特徴とする請求項２１に記載の方法。
【請求項３０】
デルタが、前記コミットを処理することができなかったスレーブ・サーバに送信されることを要求する段階をさらに含むことを特徴とする請求項２１に記載の方法。
【請求項３１】
ネットワークを通じてデータを複製する方法であって、
（ａ）前記複製を、一フェーズ法で達成すべきか、又は二フェーズ法で達成すべきかを判断し、
（ｂ）前記マスター・サーバ上に格納された前記データの変更に関し、該データの現在の状態についてのバージョン番号を含む情報のパケットを、該マスター・サーバから前記スレーブ・サーバに送信し、
スレーブ・サーバへの前記情報のパケットを受信し、
前記スレーブ・サーバ上の前記データが前記バージョン番号に対応するように更新されたかどうかを、前記スレーブ・サーバが判断することを可能にし、
前記スレーブ・サーバが前記バージョン番号に対応しない場合には、該スレーブ・サーバを更新するのに必要とされる情報を含むデルタが、前記マスター・サーバから該スレーブ・サーバに送信されることを要求する、
ことによって、一フェーズ法で達成されるように定められた複製情報を送信し、
（ｃ）前記マスター・サーバ上に格納された前記データの変更に関し、該データの現在の状態についてのバージョン番号を含む情報のパケットを、該マスター・サーバから前記スレーブ・サーバに送信し、
前記スレーブ・サーバが前記バージョン番号に対応するように更新されたかどうか、さらに、該スレーブ・サーバが前記情報のパケットを処理できるかどうかを、該スレーブ・サーバが判断することを可能にし、
前記スレーブ・サーバが更新されることを必要とするかどうか、及び該スレーブ・サーバが前記情報のパケットを処理できるかどうかを示す信号を、該スレーブ・サーバから前記マスター・サーバに送信し、
前記スレーブ・サーバが前記情報のパケットをコミットすべきであるかどうかを示す応答信号を、前記マスター・サーバから該スレーブ・サーバに送信し、

前記応答信号により前記スレーブ・サーバが前記情報のパケットをコミットすべきであることが示された場合には、該スレーブ・サーバへの該情報のパケットをコミットする、
ことによって、二フェーズ法で達成されるように定められた複製情報を送信する、
段階を含むことを特徴とする方法。

【請求項３２】

ネットワークを通じてデータを複製する方法であって、

（ａ）前記複製を、一フェーズ法で達成すべきか、又は二フェーズ法で達成すべきかを判断し、

（ｂ）前記データの現在の状態についてのバージョン番号を、マスター・サーバからスレーブ・サーバに送信し、

前記スレーブ・サーバ上のデータが前記バージョン番号に対応しない場合には、デルタが前記マスター・サーバから該スレーブ・サーバに送信されることを要求する、
ことによって、一フェーズ法で複製されるべきデータを送信し、

（ｃ）情報のパケットを前記マスター・サーバからスレーブ・サーバに送信し、

前記スレーブ・サーバが前記情報のパケットを処理できるかどうかを判断し、

前記スレーブ・サーバが前記情報のパケットを処理できる場合には、該スレーブ・サーバへの前記情報のパケットをコミットする、
ことによって、二フェーズ法で複製されるべきデータを送信する、
段階を含むことを特徴とする方法。

【請求項３３】

ネットワーク上でマスターから複数のスレーブにデータを複製する方法であって、

（ａ）前記複製を、一フェーズ法で達成すべきか、又は二フェーズ法で達成すべきかを判断し、

（ｂ）前記データの現在の状態についてのバージョン番号を前記マスターから各々のスレーブに送信し、

前記バージョン番号に対応しないデータを含むデルタが、前記マスターから各々のスレーブに送信されることを要求する、
ことによって、一フェーズ法で複製されるべきデータを送信し、

（ｃ）情報のパケットを前記マスターから各々のスレーブに送信し、

前記複数のスレーブの各々が前記情報のパケットを処理できる場合には、該スレーブへの該情報のパケットをコミットする、
ことによって、二フェーズ法で複製されるべきデータを送信する、
段階を含むことを特徴とする方法。

【請求項３４】

一フェーズ法又は二フェーズ法を用いて、ネットワーク上でマスターから複数のスレーブにデータを複製する方法であって、

（ａ）前記スレーブ上の前記データを更新するために、デルタが前記マスターから該スレーブに送信されることを、各々のスレーブが要求できるように、該データの現在の状態についてのバージョン番号を該マスターから各々のスレーブに送信することによって、一フェーズ法で複製されるべきデータを送信し、

（ｂ）あらゆるスレーブが情報のパケットをコミットできる場合、各々のスレーブによりコミットされるべき情報のパケットを前記マスターから該スレーブに送信することによって、二フェーズ法で複製されるべきデータを送信する、
段階を含むことを特徴とする方法。

【請求項３５】

一フェーズ法又は二フェーズ法を用いて、各々がクラスター・マスター及び少なくとも１つのクラスター・スレーブを含む、クラスター化されたネットワーク上でデータを複製する方法であって、

（ａ）他のクラスター・マスターがそれぞれデルタを要求できるように、前記データの

現在の状態についてのバージョン番号を第１のクラスター・マスターから他の全てのクラスター・マスターに送信することによって、一フェーズ法で複製されるべきデータを送信し、
（ｂ）前記他のクラスター・マスターが情報のパケットをコミットできる場合には、該他のクラスター・マスターによりコミットされるべき情報のパケットを、前記第１のクラスター・マスターから他のクラスター・マスターの各々に送信することによって、二フェーズ法で複製されるべきデータを送信する、
段階を含むことを特徴とする方法。
【請求項３６】
一フェーズ法により、そのクラスター・マスターを有する前記クラスターにおいて、前記データをクラスター・マスターの各々からクラスター・スレーブの各々に送信する段階をさらに含むことを特徴とする請求項３５に記載の方法。
【請求項３７】
二フェーズ法により、そのクラスター・マスターを有する前記クラスターにおいて、前記データをクラスター・マスターの各々からクラスター・スレーブの各々に送信する段階をさらに含むことを特徴とする請求項１０に記載の方法。
【請求項３８】
（ａ）前記マスター・サーバ上に格納されたデータの変更に関し、前記データの現在の状態についての現在のバージョン番号を含み、さらに、該データの前の変更及び前の変更の各々についてのバージョン番号に関する情報のパケットを、ネットワーク上でマスター・サーバから各々のスレーブ・サーバに送信する手段と、
（ｂ）前記スレーブ・サーバが前記現在のバージョン番号に対応するように更新されたかどうかを、各々のスレーブ・サーバが判断するのを可能にする手段と、
（ｃ）前記スレーブ・サーバが前の変更を逃した場合には、各々のスレーブが前記情報をコミットすることを可能にする手段と、
（ｄ）前記スレーブ・サーバが前記情報のパケットをコミットする前に、前の変更が前記マスター・サーバから該スレーブ・サーバに送信されることを、前記前の変更を逃したスレーブ・サーバの各々が要求することを可能にする手段と、
を備えることを特徴とするコンピュータ可読媒体。
【請求項３９】
ネットワークを通じてデータを複製するための、サーバ・コンピュータにより実行するためのコンピュータ・プログラム製品であって、
（ａ）前記マスター・サーバ上に格納されたデータの変更に関し、前記データの現在の状態についての現在のバージョン番号を含み、さらに、該データの前の変更及び前の変更の各々についてのバージョン番号に関する情報のパケットを、ネットワーク上でマスター・サーバから各々のスレーブ・サーバに送信するコンピュータ・コードと、
（ｂ）前記スレーブ・サーバが前記現在のバージョン番号に対応するように更新されたかどうかを、各々のスレーブ・サーバが判断するのを可能にするコンピュータ・コードと、
（ｃ）前記スレーブ・サーバが前の変更を逃した場合に、各々のスレーブ・サーバが前記情報をコミットするのを可能にするコンピュータ・コードと、
（ｄ）前記スレーブ・サーバが前記情報のパケットをコミットする前に、前の変更が前記マスターから該スレーブ・サーバに送信されるように、前記前の変更を逃したスレーブ・サーバの各々が要求することを可能にするコンピュータ・コードと、
を備えることを特徴とするコンピュータ・プログラム製品。
【請求項４０】
ネットワークを通じてデータを複製するためのシステムであって、
（ａ）前記マスター・サーバ上に格納されたデータの変更に関し、前記データの現在の状態についての現在のバージョン番号を含み、さらに、該データの前の変更及び前の変更の各々についてのバージョン番号に関する情報のパケットを、ネットワーク上でマスター

・サーバから各々のスレーブ・サーバに送信する手段と、
（ｂ）前記スレーブ・サーバが前記現在のバージョン番号に対応するように更新されたかどうかを、各々のスレーブ・サーバが判断するのを可能にする手段と、
（ｃ）前記スレーブ・サーバが前の変更を逃した場合に、各々のスレーブが前記情報をコミットすることを可能にする手段と、
（ｄ）前記スレーブ・サーバが前記情報のパケットをコミットする前に、前の変更が前記マスター・サーバから該スレーブ・サーバに送信されることを、前記前の変更を逃したスレーブ・サーバの各々が要求することを可能にする手段と、
を備えることを特徴とするシステム。
【請求項４１】
プロセッサと、
（ａ）前記マスター・サーバ上に格納されたデータの変更に関し、前記データの現在の状態についての現在のバージョン番号を含み、さらに、該データの前の変更及び前の変更の各々についてのバージョン番号に関する情報のパケットを、ネットワーク上でマスター・サーバから各々のスレーブ・サーバに送信し、
（ｂ）前記スレーブ・サーバが前記現在のバージョン番号に対応するように更新されたかどうかを、各々のスレーブ・サーバが判断するのを可能にし、
（ｃ）前記スレーブ・サーバが前の変更を逃した場合には、各々のスレーブが前記情報をコミットすることを可能にし、
（ｄ）前記スレーブ・サーバが前記情報のパケットをコミットする前に、前の変更が前記マスター・サーバから該スレーブ・サーバに送信されることを、前記前の変更を逃したスレーブ・サーバの各々が要求することを可能にする、
ように構成された、前記プロセッサにより実行されるオブジェクト・コードと、
からなることを特徴とするコンピュータ・システム。
【請求項４２】
ネットワークを通じてデータを複製するためのシステムであって、
ａ．ｉ．スタート方法を呼び出すことによりデータ複製プロセスを起動するようになり、さらに、前記データのオリジナル・コピーに関する情報を送信するようになったマスター・ユーザ層と、
ｉｉ．前記スタート方法を含み、前記マスター・ユーザ層からの前記コール及び前記データの前記オリジナル・コピーに関する前記情報を受信するようになり、さらに、該データの該オリジナル・コピーに関する該情報の少なくとも幾らかを含むデータ複製パケットを生成し送信するようになったマスター・サービス層と、
を備える、前記データのオリジナル・コピーを含むマスター・サーバと、
ｂ．ｉ．前記マスター・サービス層から前記データ複製パケットを受信し、該データ複製パケットを処理するようになり、さらに、該データ複製パケットに関する情報を送信するようになったスレーブ・サービス層と、
ｉｉ．前記データ複製パケットに関する前記情報を前記スレーブ・サービス層から受信するようになり、該情報を該データ複製パケット内に格納するようになったスレーブ・ユーザ層と、
を備える、前記マスター・サーバからの前記データのコピーを格納するようになったスレーブ・サーバと、
からなることを特徴とするシステム。
【請求項４３】
前記マスター・ユーザ層が、マスター・ユーザ及びマスター・ユーザ装置のうちの少なくとも１つと通信していることを特徴とする請求項４２に記載のシステム。
【請求項４４】
前記マスター・ユーザ層が、デルタの形態の前記データのオリジナル・コピーに関する情報を送信するようになり、前記デルタが、該データのオリジナル・コピーの前の状態と現在の状態の間の変更に関する情報を含むようになったことを特徴とする請求項４２に記

載のシステム。
【請求項４５】
前記マスター・ユーザ層が、前記データのオリジナル・コピーを更新するようになったことを特徴とする請求項４２に記載のシステム。
【請求項４６】
前記マスター・ユーザ層が、前記データのオリジナル・コピーへの変更をスレーブ・サーバ上に複製すべきでないことを示すロールバック・メッセージを送信するようになったことを特徴とする請求項４２に記載のシステム。
【請求項４７】
前記マスター・ユーザ層が、前記複製についてのタイムアウト値を設定するようになったことを特徴とする請求項４２に記載のシステム。
【請求項４８】
前記マスター・ユーザ層が、前記データのオリジナル・コピーの現在の状態と該データのオリジナル・コピーの前の状態との間のデルタを生成するようになったことを特徴とする請求項４２に記載のシステム。
【請求項４９】
前記マスター・ユーザ層が、前記データのオリジナル・コピーの現在の状態と該データの該オリジナル・コピーの前の状態との間のデルタを生成するようになったことを特徴とする請求項４２に記載のシステム。
【請求項５０】
前記マスター・ユーザ層が、前記データのオリジナル・コピーの状態の各々について固有のバージョン番号を生成するようになったことを特徴とする請求項４２に記載のシステム。
【請求項５１】
前記マスター・サービス層が、前記データ複製パケットをマルチキャストするようになったことを特徴とする請求項４２に記載のシステム。
【請求項５２】
前記マスター・サービス層が、前記データ複製パケットを一定間隔で送信するようになったことを特徴とする請求項４２に記載のシステム。
【請求項５３】
前記マスター・サービス層が、前記データ複製パケットのバージョン番号を含むようになったことを特徴とする請求項４２に記載のシステム。
【請求項５４】
前記マスター・サービス層が、前記スレーブ・サーバ上の該データのコピーを更新するのに必要な情報を前記データのオリジナル・コピーの現在の状態に含ませるようになったことを特徴とする請求項４２に記載のシステム。
【請求項５５】
さらに、前記マスター・サービス層が、デルタを含むデータ複製パケットを生成し送信するようになったことを特徴とする請求項４２に記載のシステム。
【請求項５６】
さらに、前記マスター・サービス層が、前記データのオリジナル・コピーの連続する状態の間のデルタを含むデータ複製パケットを生成し送信するようになったことを特徴とする請求項４２に記載のシステム。
【請求項５７】
さらに、前記マスター・サービス層が、前記データのオリジナル・コピーの任意の状態の間のデルタを含むデータ複製パケットを生成し送信するようになったことを特徴とする請求項４２に記載のシステム。
【請求項５８】
前記マスター・サービス層が、前記マスター・ユーザ層からデルタを要求するようになったことを特徴とする請求項４２に記載のシステム。

【請求項５９】
前記マスター・サービス層が、コミット・メッセージをスレーブ・サービス層に送信するようになったことを特徴とする請求項４２に記載のシステム。
【請求項６０】
前記マスター・サービス層が、コミット・メッセージをスレーブ・サービス層に一定間隔で送信するようになったことを特徴とする請求項４２に記載のシステム。
【請求項６１】
前記マスター・サービス層が、スレーブ・サービス層へのコミット・メッセージをマルチキャストするようになったことを特徴とする請求項４２に記載のシステム。
【請求項６２】
前記マスター・サービス層が、中止メッセージをスレーブ・サービス層に送信するようになったことを特徴とする請求項４２に記載のシステム。
【請求項６３】
前記マスター・サービス層が、中止メッセージをスレーブ・サービス層に一定間隔で送信するようになったことを特徴とする請求項４２に記載のシステム。
【請求項６４】
前記マスター・サービス層が、中止メッセージをスレーブ・サービス層にマルチキャストするようになったことを特徴とする請求項４２に記載のシステム。
【請求項６５】
前記スレーブ・ユーザ層が、スレーブ・ユーザ及びスレーブ・ユーザ装置のうちの少なくとも１つと通信していることを特徴とする請求項４２に記載のシステム。
【請求項６６】
前記スレーブ・ユーザ層が、前記スレーブ・サーバ上に格納されたデータの現在のバージョン番号を確認するようになったことを特徴とする請求項４２に記載のシステム。
【請求項６７】
前記スレーブ・ユーザ層が、前記スレーブ・サーバ上に格納されたデータへの前記データ複製パケットに関する情報をコミットするようになったことを特徴とする請求項４２に記載のシステム。
【請求項６８】
前記スレーブ・ユーザ層が、前記スレーブ・サーバ上に格納された前記データへの更新を中止するようになったことを特徴とする請求項４２に記載のシステム。
【請求項６９】
前記スレーブ・ユーザ層が、前記データ複製パケット内に含まれる準備要求を処理するようになったことを特徴とする請求項４２に記載のシステム。
【請求項７０】
前記スレーブ・ユーザ層が、前記データ複製パケット内に含まれる準備要求に関する応答を前記スレーブ・サービス層に送信するようになったことを特徴とする請求項４２に記載のシステム。
【請求項７１】
前記スレーブ・ユーザ層が、データをローカル・ディスク上に永続的にキャッシュするようになったことを特徴とする請求項４２に記載のシステム。
【請求項７２】
前記スレーブ・ユーザ層が、前記スレーブ・サーバ上の前記データのコピーの前記バージョン番号を更新するようになったことを特徴とする請求項４２に記載のシステム。
【請求項７３】
前記スレーブ・サービス層が、前記マスター・サービス層からデルタを要求するようになったことを特徴とする請求項４２に記載のシステム。
【請求項７４】
前記スレーブ・サービス層が、前記スレーブ・サービス層から、前記スレーブ・サーバ上に格納された前記データの前記現在のバージョン番号を要求するようになったことを特

徴とする請求項４２に記載のシステム。
【請求項７５】
前記スレーブ・サービス層が、コミット・メッセージを前記スレーブ・ユーザ層に送信するようになったことを特徴とする請求項４２に記載のシステム。
【請求項７６】
前記スレーブ・サービス層が、中止メッセージを前記スレーブ・ユーザ層に送信するようになったことを特徴とする請求項４２に記載のシステム。
【請求項７７】
マスター・サーバからスレーブ・サーバにデータを複製する方法であって、
マスター・サーバ上のマスター・データの現在の状態に関する情報を含むスタートコールを、該マスター・サーバ上でマスター・ユーザ・レベルからマスター・サービス・レベルに送信し、
ａ．前記スレーブ・サーバ上のスレーブ・データが現在の状態を有するかどうかを判断するために、該スレーブ・サーバ上のスレーブ・ユーザ層を確認するようになった、該スレーブ・サーバ上のスレーブ・サービス層に前記情報を送信し、
ｂ．前記スレーブ・サービス層からのデルタについての要求を、前記マスター・ユーザ層からのデルタを要求し受信するようになったマスター・サービス層に送信し、
ｃ．前記スレーブ・データを現在の状態に持って行くのに必要な前記情報を含むデルタを、前記マスター・サービス層から、前記デルタを処理し該情報を前記スレーブ・ユーザ層に送信するようになった前記スレーブ・サービス層に送信し、
ｄ．前記スレーブ・ユーザ層を用いて前記スレーブ・データを更新する、
段階を含むことを特徴とする方法。
【請求項７８】
前記マスター・ユーザ層を用いて前記データの現在の状態についてのバージョン番号を判断する段階をさらに含むことを特徴とする請求項７７に記載の方法。
【請求項７９】
マルチキャスティングにより、前記情報を前記スレーブ・サービス層に送信する段階をさらに含むことを特徴とする請求項７７に記載の方法。
【請求項８０】
前記マスター・データの現在の状態についてのバージョン番号を含む情報を、前記スレーブ・サービス層に送信する段階をさらに含むことを特徴とする請求項７７に記載の方法。
【請求項８１】
マスター・サーバからスレーブ・サーバにデータを複製する方法であって、
ａ．前記マスター・サーバ上に格納されたマスター・データにおける前の状態から現在の状態への変更に関する情報を含む新しいデルタを、マスター・サーバ上でマスター・ユーザ・レベルからマスター・サービス・レベルに送信し、
ｂ．前記スレーブ・サーバ上の前記スレーブ・データが現在の状態を有するかどうかを判断するために、該スレーブ・サーバ上のスレーブ・ユーザ層を確認するようになった該スレーブ・サーバ上のスレーブ・サービス層に、前記マスター・サービス層から前記新しいデータを送信し、
ｃ．前記スレーブ・データを前記マスター・データの前の状態に更新するのに必要な情報を含む同期デルタへの要求を、前記スレーブ・サービス層から、前記マスター・ユーザ層から同期デルタを要求し受信するようになった該マスター・サービス層に送信し、
ｄ．前記同期デルタを、前記マスター・サービス層から、前記デルタを処理し、前記情報を前記スレーブ・データにコミットされるべき前記スレーブ・ユーザ層に送信するようになった前記スレーブ・サービス層に送信し、
ｅ．前記スレーブ・ユーザ層を用いて、前記スレーブ・データへの前記新しいデルタの前記情報をコミットする、
段階を含むことを特徴とする方法。

【請求項８２】
ネットワークを通じてデータをマスター・サーバからスレーブ・サーバに複製する方法であって、
ａ．前記マスター・サーバ上の前記データのオリジナル・コピーの現在の状態に関するバージョン番号を、マスター・サービス層から、スレーブ・サービス層に送信し、
ｂ．前記スレーブ・サーバ上のデータが前記バージョン番号に対応するように更新されたかどうかを、スレーブ・ユーザ層が判断することを可能にし、
ｃ．前記スレーブ・サーバ上のデータが前記バージョン番号に対応しない場合には、デルタが前記マスター・サービス層から前記スレーブ・サービス層に送信されることを要求する、
段階を含むことを特徴とする方法。
【請求項８３】
前記スレーブ・ユーザ層が、各々のスレーブ・サーバについてローカル・ディスク上の前記データを永続的にキャッシュすることを可能にする段階をさらに含むことを特徴とする請求項７７に記載の方法。
【請求項８４】
前記マスター・ユーザ層が、前記マスター・サーバ上のデータの現在の状態についての固有のバージョン番号を判断することを可能にする段階をさらに含むことを特徴とする請求項７７に記載の方法。
【請求項８５】
スレーブ・ユーザ層がスレーブ・サーバ上の前記データを更新するのに必要な前記バージョン番号を有するデータを含ませる段階をさらに含むことを特徴とする請求項７７に記載の方法。
【請求項８６】
前記スレーブ・ユーザ層により受信されるとすぐに前記スレーブ・サーバを更新するのに必要な前記データをコミットする段階をさらに含むことを特徴とする請求項７７に記載の方法。
【請求項８７】
マスター・サーバ及び少なくとも１つのスレーブ・サーバを含む、ネットワークを通じてデータを複製する方法であって、
ａ．前記マスター・サーバ上に格納された前記データの変更に関し、前の状態についての前のバージョン番号及び該データの新しい状態についての新しいバージョン番号を含み、さらに、該データの前の変更及び前の変更の各々について前のバージョン番号に関する情報のパケットを、ネットワークでマスター・サービス層から各々のスレーブ・サーバのスレーブ・サービス層に送信し、
ｂ．前記スレーブ・サーバ上のデータが前記パケット内に含まれる前記前のバージョン番号に対応するかどうかを、各々のスレーブ・サーバのスレーブ・ユーザ層が判断することを可能にし、
ｃ．前記スレーブ・サーバ上のデータが前記パケット内に含まれる前記前のバージョン番号に対応する場合には、各々のスレーブ・ユーザ層が前記情報のパケットをコミットすることを可能にし、前記コミットはまた、前記スレーブ・サーバの前記バージョンを前記新しいバージョン番号に更新し、
ｄ．前記スレーブ・サービス層が前記情報のパケットをコミットする前に、前記スレーブを前記前のバージョン番号に更新するのに必要な前記情報を含むデルタが、前記マスター・サービス層から、そのスレーブ・ユーザ層に対応する前記スレーブ・サービス層に送信されることを、該前のバージョン番号に対応しないスレーブ・ユーザ層の各々が要求することを可能にする、
段階を含むことを特徴とする方法。
【請求項８８】
ネットワークを通じて、マスター・サーバから少なくとも１つのスレーブ・サーバにデ

ータを複製する方法であって、
ａ．前記マスター・サーバ上に格納された前記データの変更に関し、該データの現在の状態についてのバージョン番号を含む情報のパケットを、前記マスター・サーバのマスター・サービス層から、該スレーブ・サーバの前記ユーザ・サービス層に送信し、
ｂ．前記スレーブ・サーバが前記パケット内に含まれる前記バージョン番号に対応するように更新されたかどうか、さらに、該パケット内に含まれる該バージョン番号に対応するように更新することが必要な場合には、前記スレーブ・ユーザ層が前記情報のパケットを処理できるかどうかを、前記サーバの前記スレーブ・ユーザ層が判断することを可能にし、
ｃ．前記スレーブ・サーバが更新されることを必要とするかどうか、及び該スレーブ・サーバが前記更新を処理できるかどうかを示す信号を、前記スレーブ・サービス層から前記マスター・サービス層に送信し、
ｄ．前記スレーブ・ユーザ層が前記パケット内に含まれる前記情報にコミットすべきかどうかを示す応答信号を、前記マスター・サービス層から前記スレーブ・サービス層に送信し、
ｅ．前記応答信号により前記スレーブ・ユーザ層が前記パケット内に含まれる前記情報にコミットすべきであることが示された場合には、前記スレーブ・サーバへの前記情報のパケットをコミットする、
段階を含むことを特徴とする方法。
【請求項８９】
ａ．前記マスター・サーバ上の前記データのオリジナル・コピーの現在状態に関するバージョン番号を、マスター・サービス層からスレーブ・サービス層に送信する手段と、
ｂ．前記スレーブ・サーバ上のデータが前記バージョン番号に対応するように更新されたかどうかを、スレーブ・ユーザ層が判断することを可能にする手段と、
ｃ．前記スレーブ・サーバ上の前記データが前記バージョン番号に対応しない場合には、デルタが前記マスター・サービス層から前記スレーブ・サービス層に送信されることを要求する手段と、
を備えることを特徴とするコンピュータ可読媒体。
【請求項９０】
ネットワークを通じてマスター・サーバからスレーブ・サーバにデータを複製する、サーバ・コンピュータにより実行するためのコンピュータ・プログラム製品であって、
ａ．前記マスター・サーバ上の前記データのオリジナル・コピーの現在の状態に関するバージョン番号を、マスター・サービス層からスレーブ・サービス層に送信するコンピュータ・コードと、
ｂ．前記スレーブ・サーバ上の前記データが前記バージョン番号に対応するように更新されたかどうかを、スレーブ・ユーザ層が判断することを可能にするコンピュータ・コードと、
ｃ．前記スレーブ・サーバ上の前記データが前記バージョン番号に対応しない場合には、デルタが前記マスター・サービス層から前記スレーブ・サービス層に送信されることを要求するコンピュータ・コードと、
を備えることを特徴とするコンピュータ・プログラム製品。
【請求項９１】
ネットワークを通じてデータを複製するためのシステムであって、
ａ．前記マスター・サーバ上の前記データのオリジナル・コピーの現在の状態に関するバージョン番号を、マスター・サービス層からスレーブ・サービス層に送信する手段と、
ｂ．前記スレーブ・サーバ上の前記データが前記バージョン番号に対応するように更新されたかどうかを、スレーブ・ユーザ層が判断することを可能にする手段と、
ｃ．前記スレーブ・サーバ上の前記データが前記バージョン番号に対応しない場合には、デルタが前記マスター・サービス層から前記スレーブ・サービス層に送信されることを要求する手段と、

を備えることを特徴とするシステム。
【請求項９２】
プロセッサと、
（ａ）前記マスター・サーバ上に格納された前記データのオリジナル・コピーの現在の状態に関するバージョン番号を、マスター・サービス層からスレーブ・サービス層に送信し、
（ｂ）前記スレーブ・サーバ上の前記データが前記バージョン番号に対応するように更新されたかどうかを、スレーブ・ユーザ層が判断することを可能にし、
（ｃ）前記スレーブ・サーバ上の前記データが前記バージョン番号に対応しない場合には、デルタが前記マスター・サービス層から前記スレーブ・サービス層に送信されることを要求する、
ように構成された、前記プロセッサにより実行されるオブジェクト・コードと、
からなることを特徴とするコンピュータ・システム。
【発明の詳細な説明】
【発明の詳細な説明】
【０００１】
（優先権主張）
本出願は、２００１年７月１６日に出願された名称「データ複製プロトコル」の米国特許仮出願第６０／３０５，９８６号、２００１年７月１６日に出願された名称「データ複製のための階層化アーキテクチャ」の米国特許仮出願第６０／３０５，９７８号、２００１年１０月１１日に出願された名称「データ複製プロトコル」の米国特許出願第０９／９７５，５９０号、及び２００１年１０月１１日に出願された名称「データ複製のための階層化アーキテクチャ」の米国特許出願第０９／９７５，５８７号に基づく優先権を主張するものであり、これらは、出典の明示によりここで開示されたものとする。
【０００２】
（著作権の告知）
本特許文献の開示の一部は、著作権の保護を受ける内容を含んでいる。著作権者は、本特許文献が米国特許商標庁の特許ファイル又は記録において公開されたときに、何者かが特許文献又は特許の開示を複写することに異議はないが、他の場合には、如何なる場合にも全ての著作権を留保する。
【０００３】
（技術分野）
本発明は、一般に、データを転送するためのシステムに関する。より具体的には、ネットワークを通じてデータを複製するためのシステム及び方法に関する。
【０００４】
（背景技術）
分散処理システムには幾つかのタイプがある。分散処理システムは、一般に、通信媒体を通じて連結された２つのコンピュータのような複数の処理デバイスを含んでいる。あるタイプの分散処理システムは、クライアント／サーバ型ネットワークである。クライアント／サーバ型ネットワークは、少なくとも２つのデバイス、典型的には中央サーバ及びクライアントを含む。付加的なクライアントを中央サーバに連結することができ、多数のサーバがあってもよく、或いはネットワークが通信媒体を通じて連結された唯一のサーバを含んでもよい。
【０００５】
このようなネットワーク環境においては、アプリケーション又は情報を中央サーバから多数のワークステーション及び／又は他のサーバに送信することがしばしば望ましい。多くの場合、各々のワークステーションへの別々にインストールを行い、或いは、中央サーバから個々のワークステーションの各々及び／又はサーバに情報の新しいライブラリを別々に押し出す。これらの手法は時間がかかり、リソースの使用は非効率となる。各々のワークステーション又はサーバについてアプリケーションを別個にインストールすることが

、付加的かつ潜在的なエラーの源となりうる可能性もある。

【０００６】

理想的には、情報の送信は、障害に対する信頼性が高く、規模拡大が可能であるべきであり、それによってプロセスがネットワークを有効に使えるようになる。従来のソリューションでは、概して、これらの目的の一方又は両方とも達成することができない。１つの簡単な手法は、各々のスレーブに個別に連絡を取り、ＴＣＰ／ＩＰ接続のような２地点間のリンクを通じてデータを転送するマスター・サーバをもつことである。１つ又はそれ以上のスレーブが一時的に到達不能である場合、或いはスレーブが更新を処理する際にエラーが生じる場合、この手法によると、一貫性のないデータのコピーが行われることになる。これと対照的なのは複雑な分散同意プロトコルであり、これは、全てのデータのコピーが一貫していることを保証するように、スレーブの間に多くのクロストークを必要とする。

【０００７】

（発明の開示）

本発明は、ネットワーク上で達成することができるように、マスター・サーバから少なくとも１つのスレーブ又は管理対象サーバにデータを複製する方法を含む。この方法においては、複製を一フェーズ法で達成すべきか又は二フェーズ法で達成すべきかを判断することができる。複製が一フェーズ法で達成すべき場合には、マスター・サーバ上のデータの現在の状態に対応するバージョン番号を送信することができる。このバージョン番号は、ネットワーク上のあらゆるスレーブ・サーバ、又はスレーブ・サーバのサブセットにだけ送信することができる。次に、バージョン番号を受信するスレーブ・サーバは、デルタがマスターから送信されることを要求することができる。デルタは、現在のバージョン番号に対応するようにそのスレーブ上のデータを更新することが必要とされるデータを含むことができる。

【０００８】

複製が、二フェーズ法で達成されるべき場合には、マスターから各々のスレーブ、又はスレーブのサブセットに、情報のパケットを送信することができる。次に、これらのスレーブは、それらが情報のパケットをコミットできるかどうかを、マスター・サーバに応答することができる。スレーブの少なくとも幾つかがデータをコミットできる場合には、マスターは、コミットを処理すべきであるという信号をそれらのスレーブに送信することができる。コミットを処理した後、これらのスレーブは、現在のバージョン番号に更新することができる。いずれかのスレーブがコミットを処理できない場合には、コミットを中止することができる。

【０００９】

（発明を実施するための最良の形態）

本発明は、例えば、マスター・サーバ又は「管理」サーバ（「Ａｄｍｉｎサーバ」）からスレーブ・サーバの集合又は「管理対象」サーバに、データ又は他の情報の複製を提供するものである。この複製は、従来のローカル・エリア・ネットワーク又はイーサネット（登録商標）のような何らかの適切なネットワークにわたって生じることができる。１つの実施形態において、マスター・サーバは、ネットワークの全データのオリジナルの記録を所有し、このオリジナルの記録に何らかの更新が加えられることになる。更新が生じたとき、該更新と共にデータのコピーを、各々のスレーブ・サーバに伝送することができる。１つの適用例は、Ａｄｍｉｎサーバから管理対象サーバの集合に設定情報を分散することを含む。

本発明による１つのシステムにおいては、データ複製サービス（ＤＲＳ）のようなサービスが、設定及び配置情報をＡｄｍｉｎサーバから適当なドメイン内の管理対象サーバに分散させることが必要である。ユーザ・データグラム・プロトコル（「ＵＤＰ」）のようなマルチキャスト・プロトコルはフロー制御をもたず、システムを制圧することができるので、伝送制御プロトコル（「ＴＣＰ」）のような２地点間接続によって大きなデータ項目を分散させることができる。２地点間接続には、リモート・メソッド呼び出し（ＲＭＩ

）、ハイパーテキスト転送プロトコル（ＨＴＴＰ）、又は同様のプロトコルを用いることができる。

【００１０】

管理対象サーバは、ローカル・ディスク上のデータを永続的にキャッシュすることもできる。そのようなキャッシュを用いない場合には、必要なデータの転送に、許容できないほどの量の時間を必要とすることがある。転送されるべき起動データの量を減少させることによって、起動速度が増すので、管理対象サーバのキャッシュする能力は重要である。Ａｄｍｉｎサーバが到達不能の場合には、キャッシングは、起動及び／又は再起動を可能にもする。再起動はより魅力的な選択肢であり、Ａｄｍｉｎサーバがサーバに起動するよう命令する場合がそうである。しかしながら、キャッシングは、利用可能なＡｄｍｉｎサーバを用いずに、ドメインを開始する能力を提供することができる。

図１のドメイン構造１００に示されるように、１つのＡｄｍｉｎサーバ１０２と少なくとも１つの管理対象サーバ１０４が、ドメイン１０６を含むことができる。このドメイン１０６は、起動及びシャットダウンのための管理ユニットとすることができる。１つの実施形態において、ブラウザ１０８又は他のユーザ・アプリケーション又はデバイスが、Ａｄｍｉｎサーバ１０２に起動するよう伝える。次に、Ａｄｍｉｎサーバ１０２は、ドメイン１０６内の全ての管理対象サーバ１０４に起動するよう伝え、適切な設定情報を渡す。管理対象サーバ１０４が起動した後でサーバが停止した場合には、Ａｄｍｉｎサーバ１０２が利用可能であろうとなかろうと、そのサーバが自動的に再起動することが望ましい。キャッシュされたデータは、この目的のために有用なものである。

【００１１】

Ａｄｍｉｎサーバ上のデータへの更新は、バージョン間の段階的デルタとしてパッケージすることができる。デルタは、変更されるべき設定及び／又は他の情報を含むことができる。システムをオフラインにするのは望ましくないので、ドメインが動作している間に設定を更新することが好ましい。１つの実施形態において、設定の変更は、それらがＡｄｍｉｎサーバにより押し出された動的に生じる。その都度全部の設定を送信するのは不必要であり、煩わしすぎるので、設定の変更だけがデルタで送信される。

本発明によるプロトコルは、これに応じて他の適切な方法を用いることもできるが、更新の分散のための２つの方法を統合する。これらの分散方法は、一フェーズ法及び二フェーズ法と呼ぶことができ、一貫性と規模拡大可能性の間のトレードオフを提供することができる。規模拡大可能性を好む一フェーズ法において、各々のスレーブは、独自のペースで更新を獲得し処理することができる。スレーブは、異なる時間にマスターから更新を得ることができるが、更新が受信されるとすぐにデータをコミットすることができる。スレーブは、更新を処理する際にエラーに遭遇することがあるが、一フェーズ法においては、このことが、他のスレーブの更新処理を防止することはない。

【００１２】

一貫性を好む、本発明による二フェーズ法において、全てのスレーブがデータをうまく処理できるか、いずれのスレーブもデータをうまく処理できないかという点で、分散を「原子」にすることができる。中止の可能性を許容する準備段階及びコミット段階といった別個の段階がある。準備段階において、マスターは、各々のスレーブが更新を取得できるかどうかを判断することができる。スレーブが更新を受容できることを全てのスレーブが示す場合には、コミット段階において、コミットされるべきスレーブに新しいデータを送信することができる。スレーブ・サーバのうちの少なくとも１つが更新を取得できない場合には、更新を中止することができ、コミットはなされない。この場合には、管理対象サーバは、該管理対象サーバが準備をロールバックすべきであり、何も変更されないことを通知することができる。いずれの方法においても、更新がコミットされた時に到達不能なスレーブが最終的に更新を得るので、本発明によるこうしたプロトコルは信頼性があるものである。

【００１３】

本発明によるシステムはまた、一時的に利用できないサーバが最終的に全ての更新を受

信することを保証することもできる。例えば、サーバがネットワークから一時的に隔離され、その後再起動しないでネットワークに復帰する。サーバが再起動しないので、該サーバは、通常、更新を確認しない。サーバに新しい更新を定期的に確認させることによって、又はサーバが更新を受信したかどうかを調べるためにマスター・サーバに定期的に確認させることによって、ネットワークに復帰するサーバについて説明することができる。

１つの実施形態においては、マスター・サーバが、マルチキャスト「ハートビート」をスレーブ・サーバに規則的に送信する。マルチキャスト手法は信頼できないので、スレーブが任意のシーケンスのハートビートを逃すこともある。例えば、ネットワーク分割化のために、スレーブ・サーバがネットワークから一時的に切断されるか、或いはハートビートを逃させることで、スレーブ・サーバ自体が一時的にネットワークに利用できなくなることがある。したがって、ハートビートは、最近の更新についての情報のウィンドウを含むことができる。以下に説明されるように、前の更新についてのこうした情報を用いて、ネットワーク・トラフィックの量を減少させることができる。

【００１４】

各々のマスター及び各々のスレーブ内には、少なくとも２つの層、すなわち、ユーザ層及びシステム層（すなわちＤＲＳ層）がある。ユーザ層は、データ複製システムのユーザに対応することができる。ＤＲＳ層は、データ複製システム自体の実施に対応することができる。これらの参加者及び層の対話が図２に示される。

図２の起動図２００に示されるように、この実施形態におけるマスター・ユーザ層２０２及びスレーブ・ユーザ層２０４は、それぞれ、マスターＤＲＳ層２０６及びスレーブＤＲＳ層２０８内にダウンコールを行う。こうしたダウンコールは、例えば、

ｒｅｇｉｓｔｅｒＭａｓｔｅｒ（ＤＩＤ、ｖｅｒＮｕｍ、ｌｉｓｔｅｎｅｒ）

ｒｅｇｉｓｔｅｒＳｌａｖｅ（ＤＩＤ、ｖｅｒＮｕｍ、ｌｉｓｔｅｎｅｒ）

の形態をとることができ、ここで、ＤＩＤは、公知のＤＩＤの知識から得られる識別子であり、関心あるオブジェクトを指し、ｖｅｒＮｕｍは、ユーザの現在のバージョン番号としてローカルの永続的記憶装置から得られ、ｌｉｓｔｅｎｅｒは、ＤＲＳ層からのアップコールを取り扱うオブジェクトである。アップコールは、リスナ・オブジェクトについての方法を呼び出すことができる。その後、マスターは、現在のバージョン番号と共に、ハートビート又は定期的なデルタを送り始めることができる。スレーブ・ユーザ２０４から情報を得るようになったコンテナを含むことができるコンテナ層２１０が示される。可能なコンテナの例は、エンタープライズ・ジャバ・ビーンズ、ウェブ・インターフェース、及びＪ２ＥＥ（ジャバ２プラットフォーム、エンタープライズ・エディション）アプリケーションを含む。他のアプリケーション及び／又はコンポーネントは、管理クライアント２１２のようなコンテナ層２１０にプラグインすることができる。ユーザ層とＤＲＳ層の間を通信するメッセージの例は、一フェーズ法については図４に、二フェーズ法については図５に示されている。

【００１５】

図４は、本発明による階層化アーキテクチャにおける一フェーズ分散手法のために用いることのできる１つの基本プロセス４００を示す。このプロセスにおいては、マスター・ユーザ層４０２は、マスターＤＲＳ層４０６内にダウンコール４０４を行い、一フェーズ分散を開始する。このコールは、システム内の全てのスレーブに対するものとしてもよく、スレーブ・サーバのサブセットだけに対するものでもよい。このコールがサブセットに対するものである場合、マスター・ユーザ層４０２は、更新の範囲、或いはどのスレーブが更新を受信すべきであるかを判断することができる。

【００１６】

マスターＤＲＳ層は、マスター上のデータの現在のバージョン番号を含むハートビート４０８をスレーブＤＲＳ層にマルチキャストし始める４１０。スレーブＤＲＳ層４１０は、スレーブ・ユーザ層４１４からそのスレーブについての現在のバージョン番号を要求する４１２。次に、スレーブ・ユーザ層４１４は、そのスレーブのバージョン番号を有するスレーブＤＲＳ層４１０に応答する４１６。スレーブが同期しているか、又は既に現在の

バージョン番号をもっている場合には、次の更新までそれ以上の要求はなされない。スレーブが同期しておらず、該スレーブが更新の範囲内にある場合には、スレーブＤＲＳ層４１０は、該スレーブをマスター上のデータの現在のバージョン番号に更新するために、マスターＤＲＳ層４０６からデルタを要求することができる４２０。マスターＤＲＳ層４０６は、マスター・ユーザ層４０２が、スレーブを更新するためにデルタを生成することを要求する４２２。次に、マスター・ユーザ層４０２は、デルタをマスターＤＲＳ層４０６に送信４２４し、該マスターＤＲＳ層は、該デルタ及びマスターの現在のバージョン番号をスレーブＤＲＳ層４１０に転送し４２６、該スレーブＤＲＳ層は、デルタをコミットされるべきスレーブ・ユーザにデルタを送信する４２８。ハートビート４０８がスレーブにより受信されてからマスターが更新された場合は、現在のバージョン番号がデルタと共に送信される。

【００１７】

マスターＤＲＳ層４０６は、バージョン番号を含むマルチキャスト・ハートビートを、スレーブ・サーバに定期的に送信し続けることができる４０６。このことは、利用できなかった、又はデルタを受信し処理することができなかった如何なるサーバも、該サーバがデータの現在のバージョン番号を有していないことを判断し、後にスレーブがシステムに復帰した時などにデルタを要求することを可能にする４２０。

図５は、本発明による階層化アーキテクチャにおける二フェーズ分散手法のために用いることのできる１つの基本プロセス５００を示す。このプロセスにおいては、マスター・ユーザ層５０４は、マスターＤＲＳ層５０６内にダウンコールを行い、二フェーズ分散を開始する５０４。マスター・ユーザ層５０２は、再び更新の範囲を判断することを必要とし、更新処理のための「タイムアウト」値を設定することができる。

【００１８】

マスターＤＲＳ層５０６は、新しいデルタをスレーブＤＲＳ層５１０に送信する５０８。スレーブＤＲＳ層５１０は、新しいデルタについて、準備要求をスレーブ・ユーザ層５１４に送信する５１２。次に、スレーブ・ユーザ層５１４は、スレーブが新しいデルタを処理できようとできまいとスレーブＤＲＳ層５１０に応答する５１６。スレーブＤＲＳ層は、応答をマスターＤＲＳ層５０６に転送する５１８。スレーブが同期していないために要求を処理できない場合には、マスターＤＲＳ層５０６は、マスター・ユーザ層５０２にアップコールを行い、スレーブを同期させるデルタを生成し、該デルタをコミットする５２０。マスター・ユーザ層５０２は、同期デルタをマスターＤＲＳ層に送信し５２２、該マスターＤＲＳ層は、同期デルタをスレーブＤＲＳ層５１０に転送する５２４。スレーブが同期デルタを処理できる場合には、スレーブＤＲＳ層５１０は、同期応答をマスターＤＲＳ層５０６に送信し５２６、ここで該スレーブが新しいデルタを処理できるようになる。スレーブが同期デルタを処理できない場合には、スレーブＤＲＳ層５１０は、適切な同期応答をマスターＤＲＳ層５０６に送信する５２６。次に、マスターＤＲＳ層５０６は、スレーブが新しいデルタを処理できると該スレーブ・サーバが応答したかどうかによって、コミット・メッセージ又は中止メッセージをスレーブＤＲＳ層５１０に一定間隔で送信する５２８。例えば、全てのスレーブがデルタを準備できる場合には、マスターは、コミット信号を一定間隔で送ることができる。その他の場合には、マスターは、コミット信号を一定間隔で送信することができる。ハートビートは、更新の範囲も含むので、スレーブは、該ハートビート内に含まれる情報を処理すべきかどうかを知るようになる。

【００１９】

スレーブＤＲＳ層は、このコマンドをスレーブ・ユーザ層５１４に転送し５３０、次いで、該スレーブ・ユーザ層は、新しいデルタの更新をコミットするか又は中止する。準備段階がマスター・ユーザ層５０２によって設定されたタイムアウト値内に完了しなかった場合には、マスターＤＲＳ層５０６は、中止を全てのスレーブに自動的に一定間隔で送信することができる５２８。このことは、例えば、マスターＤＲＳ層５０６がスレーブのうちの少なくとも１つに連絡し、該スレーブがコミットを処理できるかどうか判断することかできない時に生じる。タイムアウト値は、マスターＤＲＳ層５０６が、更新を中止する

前に指定の時間スレーブに連絡をとろうとすることができるように、設定することができる。

【００２０】

一フェーズ法における更新の場合、これらのハートビートは、各々のスレーブに、該スレーブのデータの現在のバージョンから始まるデルタを要求させることできる。こうしたプロセスは、図６のフローチャートに示されている。本発明による階層化アーキテクチャを利用してもしなくともよいこの基本プロセス６００において、マスター・サーバ上の現在のデータのバージョン番号は、マスター・サーバからスレーブ・サーバに送信される６０２。スレーブ・サーバは、該スレーブ・サーバが現在のバージョン番号に更新されたかどうかを判断する６０４。スレーブが現在のバージョンでない場合には、スレーブ・サーバを更新するために、デルタがマスター・サーバから送信されることを要求する６０６。デルタがスレーブ・サーバに送信された時には、スレーブ・サーバは、スレーブ・データを現在のバージョンに更新する６０８。次に、スレーブ・サーバは、バージョン番号を現在のバージョン番号に更新する６１０。

【００２１】

二フェーズ法における更新の場合、マスターは、該マスターが直前のバージョンからのデルタを積極的に各々のスレーブに送信する準備段階で始めることができる。こうしたプロセスが、図７のフローチャートに示されている。本発明による階層化アーキテクチャを利用してもしなくともよいこの基本プロセス７００においては、情報のパケットが、マスターからスレーブ・サーバに送信される７０２。パケットを受信する各々のスレーブ・サーバは、該スレーブ・サーバがそのパケットを処理し、現在のバージョンに更新できるかどうか判断する７０４。スレーブ・サーバがパケットを処理できるかどうかを示す、パケット応答を受信するスレーブ・サーバの各々は、マスター・サーバに応答する７０６。全てのスレーブ（これにデルタが送信される）が、あるタイムアウト時間内にデルタの処理に成功したことを確認した場合には、マスターは、更新をコミットするように決定することができる。他の場合には、マスター・サーバは、更新を中止するように決定することができる。一旦決定が下されると、マスター・サーバは、更新をコミットするべきか中止すべきかを示すメッセージをスレーブ・サーバに送信する７０８。決定がコミットすべきである場合には、各々のサーバは、コミットを処理する７１０。スレーブの中の１つによりコマンドを逃した場合には、ハートビートを更に用いて、コミットが生じたか、或いは中止が生じたかを知らせることができる。

【００２２】

スレーブは、最初にマスターから現在のバージョン番号を得ることなく、キャッシュされたデータを用いて、直ちに起動及び／又は再起動するように構成することができる。上述したように、本発明による１つのプロトコルは、スレーブがローカル・ディスク上のデータを永続的にキャッシュすることを可能にする。このキャッシングは、転送されることが必要なデータの量を減少させることによって、システムの起動に必要とされる時間を減少させ、規模拡大可能性を改善するものである。このプロトコルは、マスターが到達不能な場合に、スレーブが起動及び／又は再起動することを可能にすることにより、信頼性を改善することができ、さらに、バージョン間の段階的デルタとして更新をパッケージすることを可能にすることができる。キャッシュ・データが存在しない場合には、スレーブは、マスターを待つか、又はデータ自体を取り出すことができる。スレーブがキャッシュを有する場合には、該スレーブは、依然として同期しないで起動することを望まないであろう。スレーブが待つことを承知している場合には、起動時間を減少させることができる。

【００２３】

このプロトコルは、状況に合わせて、マスター又はスレーブがデータ転送のイニシアチブを取ることができるという点で、双方向性とすることができる。例えば、スレーブは、ドメインの起動中、マスターからデルタを取り出すことができる。スレーブが、デルタを更新しようと意図するものと異なるバージョンにあると判断した時には、該スレーブは、デルタがその現在のバージョンからの現在のシステム・バージョンに要求することができ

る。スレーブはまた、一フェーズ分散中にデルタを取り出すこともできる。ここで、システムは、ハートビートを読み取り、更新をとらえ損なったことを判断し、適切なデルタを要求することができる。

スレーブはまた、例外的な状況から回復するのに必要とされる時にデルタを取り出すこともできる。例外的な状況は、例えば、システムのコンポーネントが同期していない時に存在する。スレーブかデルタを取り出す時、該デルタは、データの任意のバージョン間にあることができる。換言すれば、どんなに多くの隔たった反復がそれらのバージョン間にあっても、デルタは、スレーブの現在のバージョンとシステム（又はドメイン）の現在のバージョンの間にあることができる。この実施形態においては、ハートビートの可用性、及びデルタを受信する能力が、システムの同期をもたらすことができる。

【００２４】

スレーブがデルタを取り出す能力に加えて、マスターは、二フェーズ分散中、デルタをスレーブに押し出す能力を有することができる。１つの実施形態において、これらのデルタは、常に、データの連続するバージョンの間にある。この二フェーズ分散手法は、参加者の間に不一致が生じる可能性を最小にすることができる。スレーブ・ユーザは、更新をクライアントに公開することなく、又は更新をロールバック不能にすることなく、できる限り速く準備を処理することができる。このことは、競合についてサーバを確認するようなタスクを含むことができる。「ディスクが満杯である」又は「一貫性のない設定である」といったメッセージを送信することによって、スレーブのいずれかがエラーを知らせる場合には、更新を全体的にロールバックすることができる。

【００２５】

しかしながら、非一貫性が生じる可能性が依然としてある。例えば、ソケットを開くことができないといった理由のために、コミットを処理する際にエラーが生じ得る。サーバはまた、様々な時間に更新をコミットし、公開することができる。データが、全く同時に全ての管理対象サーバに到達するはできないので、幾らかのリップル効果が生じ得る。このリップル効果を最小にしようとして、マルチキャストを使用することが小さな時間窓を提供することができる。１つの実施形態においては、コミットをとらえ損なった場合に準備されたスレーブが停止する、信号をとらえ損なった場合にマスターがクラッシュするなど。

マルチキャストに対するベストエフォート型手法は、スレーブ・サーバにコミット信号を逃させることがある。コミット段階中マスターが途中でクラッシュした場合、回復のためのログイン又は手段はないであろう。マスターが、残りのスレーブに、該マスターがコミットする必要があることを伝える方法はない。中止する際、バージョンが適切にロールバックされない場合には、幾つかのスレーブが、データをコミットすることで終わることができる。１つの実施形態において、残りのスレーブは、一フェーズ分散を用いて更新を得ることができる。このことは、例えば、管理対象サーバが、Ａｄｍｉｎサーバから受信したハートビートに応答してデルタを取り出す時に起こり得る。この手法はシステムの規模拡大可能性を維持し、如何なるコミット・エラー又はバージョン・エラーをも回避するためにシステムが分散を制限した場合に失われるであろう。

【００２６】

システムにより管理される各々のデータ項目は、ドメイン全体にわたって公知である、固有の長寿命ドメイン識別子（ＤＩＤ）をもつように構造化することができる。データ項目は、各々がドメイン内のサーバの幾つかのサブセットに関連する多くのコンポーネントからなる、大きく複雑なオブジェクトとすることができる。これらのオブジェクトは、一貫性のあるユニットとすることができるので、幾つかの小さなオブジェクトより、少数の大きなオブジェクトをもつ方が望ましい。例として、単一のデータ項目又はオブジェクトは、ｃｏｎｆｉｇ．ｘｍｌファイル又はアプリケーション－ＥＡＲファイルといったコードファイルを含むシステムについての全ての設定情報を表すことができる。データ項目内の所定のコンポーネントは、スレッドの数に関して個々のサーバに関連することができ、配置されたサービスに関してクラスターに関連することができ、或いは安全証明書に関す

るドメイン全体に関連することもできる。２つのバージョンの間のデルタは、これらのコンポーネントのうちの幾つか又は全てについての新たな値から構成することができる。例えば、コンポーネントは、ドメインのメンバーに配置された全てのエンタープライズ・ジャバ・ビーンズを含むことができる。デルタは、これらのジャバ・ビーンズのサブセットだけに対する変更を含むことができる。

【００２７】

デルタの「範囲」は、該デルタ内の関連するコンポーネントを有する全てのサーバの組を指すことができる。本発明によるＡｄｍｉｎサーバは、デルタの範囲を判断するために、設定の変更を解釈することができる。マスター上のＤＲＳシステムは、データを適切なスレーブに送信するために、その範囲を知る必要がある。マスターが、各々の更新においてサーバのサブセットに触れるだけでよい時、あらゆる設定の更新をあらゆるサーバに送信することは、時間とリソースの無駄であろう。

分散を制御するために、マスター・ユーザは、連続するバージョン間のデルタと共に各々の更新の範囲を提供することができる。この範囲は、ドメイン内の同じネームスペースからとることのできる、サーバ及び／又はクラスターに関する名前の組として表すことができる。１つの実施形態においては、ＤＲＳは、名前をアドレスにマッピングするために、リゾルバ・モジュールを用いる。クラスター名は、そのクラスター内の全てのサーバのアドレスの組にマッピングすることができる。これらのアドレスは、仮想マシンに関するように、相対的なものとすることができる。当該技術分野において周知であり使われているように、リゾルバは、介在するファイアウォールがあるかどうかを判断し、サーバが「ファイアウォールの内側」にあるかどうかに関する、「内側」アドレス又は「外側」アドレスのいずれかを返すことができる。Ａｄｍｉｎサーバ又は他のサーバは、設定データを用いて対応するリゾルバを初期化することができる。

【００２８】

管理されるデータ項目の各々についての固有の長寿命ドメイン識別子（ＤＩＤ）と共に、データ項目の各々のバージョンは、長寿命のバージョン番号をもつこともできる。適切なバージョンに関する混乱のために、サーバが不適正に更新したり更新に失敗したりしないように、各々のバージョン番号を更新の試みに固有のものとすることができる。同様に、中止された二フェーズ分散のバージョン番号を再使用することはできない。マスターは、バージョン番号だけが与えられた任意の２つのバージョン間のデルタを生成することができる。マスターが、こうしたデルタを生成できない場合には、データ又はアプリケーションの完全なコピーを提供することができる。

【００２９】

データ複製サービスは、できる限り一般的なものにしておくことが望ましい。したがって、このシステムのユーザに少数の仮定を課すことができる。このシステムは、例えば、

・このシステムは、バージョン番号を増加させる方法を含むことができる、

・このシステムは、バージョン番号をマスターにもスレーブにも永続的に格納することができる、

・このシステムは、バージョン番号を比較し、等しいかどうかを判断する方法を含むことができる、

という３つの主な仮定に依存することができる。これらの仮定は、インターフェース「ＶｅｒｓｉｏｎＮｕｍｂｅｒ」のような、ＤＲＳインターフェースのユーザレベルの実施によって提供することができる。こうしたインターフェースは、ユーザが、バージョン番号抽象化の特定の概念及び実施を提供することを可能にできる一方、システムがそのバージョン番号の属性にアクセスを有することを保証する。例えば、ジャバでは、ＶｅｒｓｉｏｎＮｕｍｂｅｒインターフェースを、以下のように実行することができる。

```
package weblogic.drs;
public interface VersionNumber extends Serializable {
  VersionNumber increment();
  void persist() throws Exception;
  boolean equals (VersioNumber anotherVN);
  boolean strictlyGreaterThan(VersionNumber anotherVN);
}
```

ユーザがシステムに提供できるこの抽象化の単純化した実施は、大きな正の整数となるであろう。この実施はまた、当該技術分野において「直列化可能」と呼ばれる、システムがネットワークを介してマスターからスレーブまでデルタ情報を転送できることをも保証することができる。

【００３０】

上記の抽象化を用いる場合、ユーザレベルにおけるデルタの詳細な内容の概念から抽象化することが有用である。このシステムは、デルタ情報構造の知識を必要としなくてもよく、実際に、該構造を判断することさえできなくてもかまわない。デルタの実施を、直列化可能とすることもでき、システムがネットワークを介してマスターからスレーブまでデルタのバージョン情報を転送できることを保証する。

適切なＤＩＤ及びバージョン番号と共に、各々のデータ項目についての記録のコピーをマスターに永続的に格納させることが望ましい。二フェーズ分散を始める前に、マスターは、提案された新しいバージョン番号を永続的に格納し、マスターが機能しなくなる場合に該新しいバージョン番号が再使用されないことを保証することができる。スレーブは、そのＤＩＤ及びバージョン番号と共に、各々の関連データ項目の最新のコピーを永続的に格納することができる。スレーブがその都度データ又はプロトコルを取得しなければならないように、必要なキャッシングを行うように該スレーブを構成することもできる。このことは、あらゆる場合について望ましいわけではないが、起こり得る特定の状況を取り扱うために許容することができる。

【００３１】

さらに、本発明によるシステムは、同時制限を含むことができる。例えば、所定の範囲にわたる所定のＤＩＤについての更新の二フェーズ分散中、特定の操作が許容されないようにすることができる。こうした操作は、空でない交差部分を有する範囲にわたって同じＤＩＤ上の範囲のメンバーシップを修正するといった、一フェーズ法の更新又は二フェーズ法の更新を含むことができる。

少なくとも１つの実施形態において、マスターＤＲＳは、ドメイン内の各々のサーバのスレーブＤＲＳに対して、ハートビート又はパケット情報を規則的にマルチキャストする。各々のＤＩＤについて、ハートビートは、各々の更新バージョン番号、前のバージョンに対するデルタの範囲、及びその更新がコミットされたか又は中止されたかを含む、最新の更新についての情報ウィンドウを含むことができる。現在のバージョンのついての情報をいつも含ませることができる。正確さ又はも有効性のためではなく、マスターに戻るトラフィックの量を最小にするために、古いバージョンについての情報を用いることもできる。

【００３２】

デルタ内に古いバージョン情報を含ませる場合、スレーブは、準備時に予期していた更新の一部分をコミットし、より新しい更新を取り扱うために新しいデルタを要求することができる。矢継ぎ早の更新がウィンドウを許容できない寸法まで増加させることがあるが、少なくとも幾つかの一定の設定可能な数のハートビートの間、所定のバージョンについての情報を含ませることができる。別の実施形態においては、一旦マスターが全てのスレーブが更新を受信したと判断すると、古いバージョンについての情報を廃棄することができる。

マルチキャスト・ハートビートは、考慮すべき幾つかの特性をもつ。これらのハートビートは、非同期式すなわち「一方通行」とすることができる。その結果、スレーブがハートビートに応答するまでには、マスターは、新たな状態に進んでいることがある。さらに、全てのスレーブが正確に同時に応答するとは限らない。したがって、マスターは、スレーブがその状態についての知識をもっていないと仮定し、デルタが更新することを意図するものを含むことができる。スレーブがハートビートの任意のシーケンスを逃すこともあるので、これらのハートビートも、信頼性を欠くものである。このことは、ハートビート内に古いバージョン情報を再び含ませることにつながる。１つの実施形態においては、ハートビートは、それらが送信された順序でスレーブに受信される。例えば、スレーブは、該スレーブがバージョン６をコミットするまで、バージョン７をコミットすることができない。サーバは、該サーバが６を受信するまで待つか、或いは単純に６を廃棄し、７をコミットすることができる。この順序付けは、バージョンが逆戻りすることによりもたらされる混乱の可能性を排除することができる。

【００３３】

上述のように、ドメインは、図３に示されるようなクラスター化を利用することもできる（マルチキャスト・ハートビート・スライドの特性）。この実施形態についての一般的なネットワーク・トポロジーは、マスターを含むハブ・アイランドに接続されたマルチキャスト・アイランドの集合である。マルチキャスト・トラフィックは、ハブから外へ２地点間で転送される。一フェーズ法で分散することができる小さなデルタを、マルチキャストを通じて直接伝送することができる。他の全ての場合、デルタは、２地点間リンクを通じて伝送することができる。ツリー構造化された２地点間転送スキームをハブ・スポーク式マルチキャスト構造の上にオーバーレイし、マスターにおけるボトルネックを減少させることができる。

【００３４】

図３のドメイン図３００において、管理対象サーバ３０２の１つ又はそれ以上を、クラスター３０４とも呼ばれるマルチキャスト・アイランド内にまとめることができる。ドメイン３０８のＡｄｍｉｎサーバ３０６は、ハブ・アイランド３１２のマスターとして働き、ブラウザ３１０を介するといった、ドメインへの入力点である。Ａｄｍｉｎサーバ３０６は、クラスター・マスターと呼ばれる、クラスター内の管理対象サーバの１つに連絡する。この実施形態におけるＡｄｍｉｎサーバは、デルタ又はメッセージを各々のクラスター・マスターにマルチキャストすることができ、次に、各々のクラスター・マスターは、マルチキャストによってデルタ又はメッセージをそのクラスター内の他の管理対象サーバに転送する。クラスター・マスターは、如何なる設定情報も所有せず、代わりに、Ａｄｍｉｎサーバから情報を受信する。クラスター・マスターがオフラインになるか又はクラッシュした場合には、ドメイン内の別の管理対象サーバがクラスター・マスターにとって代わることができる。この場合、オフラインのサーバが第２のクラスター・マスターとして復帰することを防止するために、機構を所定の位置に置くことができる。このことは、クラスター又はシステム・インフラストラクチャにより取り扱うことができる。

【００３５】

２つ以上のドメインが存在することもできる。この場合は、ネスト状ドメインすなわち「シンジケート」をおくことができる。各々のドメイン・マスターが他のドメイン・マスターに情報を押し出す能力をもつことができるので、各々のドメイン・マスターに直接触れることにより、情報を該ドメイン・マスターに行き渡らせることができる。しかしながら、ドメイン・マスターにマルチキャストすることは望ましくない。

一フェーズ分散において、マスター・ユーザは、更新の分散をトリガするために、ダウンコールを行うことができる。こうしたダウンコールは、

ｓｔａｒｔＯｎｅＰｈａｓｅ（ＤＩＤ，ｎｅｗＶｅｒＮｕｍ，ｓｃｏｐｅ）

の形態をとることができ、ここで、ＤＩＤは、更新されたデータ項目又はオブジェクトのＩＤであり、ｎｅｗＶｅｒＮｕｍは、そのオブジェクトの新しいバージョン番号であり、ｓｃｏｐｅは、その更新が適用される範囲である。マスターＤＲＳは、新しいバージョン

番号を進ませ、該新しいバージョン番号をディスクに書き込み、その後のハートビート内に情報を含ませることにより、応答することができる。

【００３６】

スレーブＤＲＳがハートビートを受信した時、該スレーブＤＲＳは、該スレーブＤＲＳが関心ある最新の更新に関する情報のウィンドウを分析することにより、取り出しを必要とするかどうかを判断することができる。スレーブの現在のバージョン番号がウィンドウ内にあり、該スレーブが後にコミットされる更新のいずれの範囲にも入っていない場合には、如何なるデータも取り出すことなく、単純に最新のバージョン番号に進ませることができる。このプロセスは、スレーブが最新である些細な場合を含むことができる。他の場合には、スレーブＤＲＳは、マスターＤＲＳからのデルタに対する２地点間コール、又は、

ｃｒｅａｔｅＤｅｌｔａ（ＤＩＤ，ｃｕｒＶｅｒＮｕｍ）

の形態をとることができる別の同様の要求を行うことができ、ここで、ｃｕｒＶｅｒＮｕｍは、スレーブの現在のバージョン番号であり、ドメイン・マスター又はクラスター・マスターに送り返される。この要求を取り扱うために、マスターＤＲＳは、ｃｒｅａｔｅＤｅｌｔａ（ｃｕｒＶｅｒＮｕｍ）のようなアップコールを行うことができる。このアップコールは、デルタ及び新しいバージョン番号を獲得し、それらをスレーブＤＲＳに戻すために、適切なリスナを通じて行うことができる。スレーブが最後にハートビートを受信してから新しいバージョン番号が変更されたかもしれないので、新しいバージョン番号を含ませるべきである。デルタは、最も新しくコミットされた更新までのものだけにすることができる。如何なる進行中の二フェーズ更新も、別個の機構によって取り扱うことができる。次に、スレーブＤＲＳは、ｃｏｍｍｉｔＯｎｅＰｈａｓｅ（ｎｅｗＶｅｒＮｕｍ，ｄｅｌｔａ）のような、スレーブ・ユーザへのアップコールを行い、その後、新しいバージョン番号に進むことができる。

【００３７】

二フェーズ更新分散をトリガするために、マスター・ユーザは、ｓｔａｒｔＴｗｏＰｈａｓｅ（ＤＩＤ、ｏｌｄＶｅｒＮｕｍ、ｎｅｗＶｅｒＮｕｍ、ｄｅｌｔａ、ｓｃｏｐｅ、ｔｉｍｅｏｕｔ）のようなダウンコールを行うことができ、ここで、ＤＩＤは更新されるべきデータ項目又はオブジェクトのＩＤであり、ｏｌｄＶｅｒＮｕｍは前のバージョン番号であり、ｎｅｗＶｅｒＮｕｍは新しいバージョン番号であり（前のバージョン番号からの１段階）、ｄｅｌｔａ取り出されるべき連続するバージョン間のデルタであり、ｓｃｏｐｅは更新の範囲であり、ｔｉｍｅｏｕｔはそのジョブについての最大有効期間である。「準備」及び「コミット」は同期式であるので、ジョブに対して特定の期限を設定することが望ましい。前のバージョン番号を含ませることができ、これに対して異なるバージョン番号のサーバはデルタを取らない。

【００３８】

１つの実施形態におけるマスターＤＲＳは、範囲内にある全てのサーバを検討し、各々のスレーブＤＲＳに対して、ｐｒｅｐａｒｅＴｗｏＰｈａｓｅ（ＤＩＤ、ｏｌｄＶｅｒＮｕｍ、ｎｅｗＶｅｒＮｕｍ、ｄｅｌｔａ、ｔｉｍｅｏｕｔ）のような２地点間コールを行う。次に、スレーブは、適切なタイムアウト値を取得することができる。バイナリ・コードを含むデルタのように、デルタが大きい場合には、２地点間プロトコルを用いることができる。一フェーズ法を用いて、例えば、キャッシュ・サイズの修正のような小さな設定変更だけを含むことができる小さな更新を行うことができる。アプリケーションの追加のような大きな変更が一貫性をもってサーバに到達することがより重要なので、この手法を用いることができる。代わりに、マスターが、存在する場合にはクラスター・マスターのところに行き、該クラスター・マスターにコールを行わせることができる。クラスター・マスターの代理を務めるマスターをもつことは、システムの規模拡大可能性を改善することができる。

【００３９】

１つの実施形態において、スレーブ又はクラスター・マスターへの各々のコールは、マ

スターＤＲＳにより取り扱われる「到達不能」、「ＯｕｔＯｆＳｙｎｃ」、「Ｎａｋ」、「Ａｃｋ」といった４つの応答のうちの１つを生成する。応答が「到達不能」である場合には、当該サーバに到達することができず、再試行のためにキューに入れることができる。応答が「ＯｕｔｏｆＳｙｎｃ」である場合には、再試行のためにキューに入れることができる。その間、サーバは、マスターからの取り出しを用いて該サーバ自体を同期させようとするので、再試行時にデルタを受信できるようになる。応答が「ＮｏＡｃｋ」すなわち肯定応答がない場合には、ジョブは中止される。サーバがジョブを許容できない場合にこの応答が与えられる。応答が「Ａｃｋ」である場合には、如何なる動作もとられない。

【００４０】

スレーブを準備するため、マスターＤＲＳは、ｐｒｅｐａｒｅＴｗｏＰｈａｓｅのような方法を呼び出すことができる。マスターＤＲＳからの「準備」要求を受信すると、スレーブＤＲＳは、まず、その現在のバージョン番号が更新されるべき古いバージョン番号と等しいかどうか確認する。等しくない場合には、スレーブは、「ＯｕｔＯｆＳｙｎｃ」応答を返すことができる。次に、スレーブは、あたかもちょうどハートビートを受信したかのように、マスターＤＲＳからデルタを取り出すことができる。最終的に、マスターＤＲＳは、ｐｒｅｐａｒｅＴｗｏＰｈａｓｅを再試行することができる。この手法は、マスターにデルタを取り出させるよりも簡単であるが、マスターの慎重な設定を必要とする。応答を長く待ちすぎることはジョブのタイムアウトを招くので、マスターの設定が必要となる。さらに、十分に応答を待たないことは、「ＯｕｔＯｆＳｙｎｃ」応答を得るのに余分な要求をもたらすことになる。スレーブからの取り出し要求の完了時に、再試行をトリガすることが好ましい。

【００４１】

スレーブが同期している場合には、該スレーブは、ｐｒｅｐａｒｅＴｗｏＰｈａｓｅ（ｎｅｗＶｅｒＮｕｍ，ｄｅｌｔａ）のように、できる限り深くサーバ内のスレーブ側のクライアント層に、アップコールを行うことができる。次に、返される結果物としての「Ａｃｋ」又は「Ｎａｋ」を、マスターＤＲＳに送信することができる。応答が「Ａｃｋ」であった場合には、スレーブは、特別の準備状態になることができる。応答が「Ｎａｋ」であった場合には、スレーブは、如何なる更新の記録もフラッシュすることができる。何らかの理由で、後でコミットされることになる場合には、スレーブは、これを一フェーズ分散として獲得することができ、その後機能しなくなる。

【００４２】

マスターＤＲＳがタイムアウト期間内にあらゆるサーバからなんとか「Ａｃｋ」を収集する場合には、該マスターＤＲＳは、ｔｗｏＰｈａｓｅＳｕｃｃｅｅｄｅｄ（ｎｅｗＶｅｒＮｕｍ）のようなコミット・アップコールを行い、新しいバージョン番号に進むことができる。マスターＤＲＳがいずれかのサーバから「Ｎａｋ」を受信した場合、又はタイムアウト時間が終了した場合には、マスターＤＲＳは、ｔｗｏＰｈａｓｅＦａｉｌｅｄ（ｎｅｗＶｅｒＮｕｍ，ｒｅａｓｏｎ）のような中止アップコールを行い、バージョン番号を変更しないままにしておくことができる。ここで、ｒｅａｓｏｎは、如何なる「Ｎａｋ」応答のロールアップも含む例外である。いずれにしても、中止／コミット情報をその後のハートビートに含ませることができる。

どんな時でも、マスターＤＲＳは、ｃａｎｃｅｌＴｗｏＰｈａｓｅ（ｎｅｗＶｅｒＮｕｍ）のような取消しダウンコールを行うことができる。次に、マスターＤＲＳは、ジョブが進行中でない場合には例外を廃棄することによりこのコールを取り扱うことができ、或いは停止が生じる予定であるように働く。

【００４３】

準備されたスレーブＤＲＳが、新しいバージョン番号がコミットされたことを示すハートビートを取得した場合には、該スレーブＤＲＳは、ｃｏｍｍｉｔＴｗｏＰｈａｓｅ（ｎｅｗＶｅｒＮｕｍ）のようなアップコールを行い、新しいバージョン番号に進むことができる。準備されたスレーブＤＲＳが、代わりに、新しいバージョン番号が中止されたことを示すハートビートを取得した場合には、該スレーブはジョブを中止することができる。

ウィンドウが新しいバージョン以上に進んだ場合にスレーブがハートビートを取得したとき、該スレーブは、同じデータ項目について新たなｐｒｅｐａｒｅＴｗｏＰｈａｓｅコールを取得するか、又はジョブをタイムアウトする。このような場合には、スレーブは、ａｂｏｒｔＴｗｏＰｈａｓｅ（ｎｅｗＶｅｒＮｕｍ）のようなアップコールを行い、バージョン番号を変更しないままにしておくことができる。このことは、スレーブが準備された後、該スレーブがコミットする前に、マスター・サーバが機能しなくなる場合のような、状況を適切に取り扱うことを保証する１つの方法である。

【００４４】

本発明の好ましい実施形態の上記の説明は、図示及び説明目的で提供された。これは、網羅的であるか又は本発明を開示された精密な形態に制限することを意図するものではない。明らかに、多くの修正及び変形が当業者には明らかであろう。実施形態は、本発明の原理及びその実際の用途を最もよく説明するために選択されて述べられたものであり、したがって、当業者が種々の実施形態について及び考慮される特定の用途に適当な種々の修正をもって本発明を理解できるようにするものである。本発明の範囲は、特許請求の範囲及びその均等技術により定義されることが意図される。

【図面の簡単な説明】

【００４５】

【図１】本発明の１つの実施形態によるドメイン構造の図である。

【図２】本発明の１つの実施形態による階層化アーキテクチャの図である。

【図３】本発明の１つの実施形態によるクラスター化されたドメイン構造の図である。

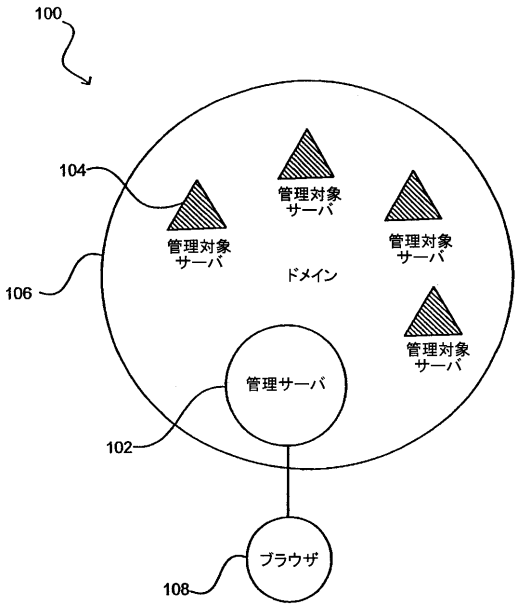
【図４】本発明の１つの実施形態による階層化アーキテクチャのための一フェーズ法の図である。

【図５】本発明の１つの実施形態による階層化アーキテクチャのための二フェーズ法の図である。

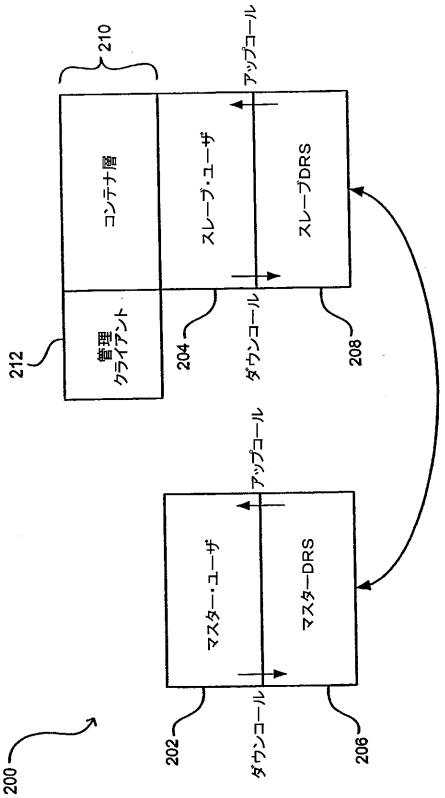
【図６】本発明の１つの実施形態による一フェーズ法のフローチャートである。

【図７】本発明の１つの実施形態による二フェーズ法のフローチャートである。

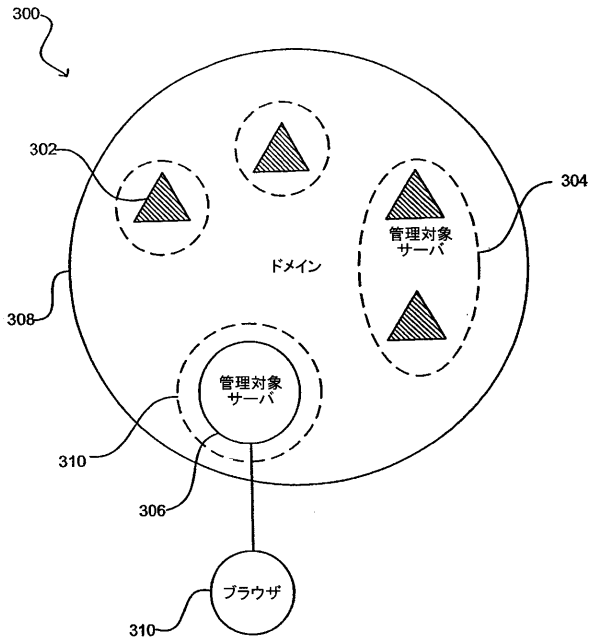
【図１】


100
104
管理対象サーバ
管理対象サーバ
管理対象サーバ
106
ドメイン
管理対象サーバ
管理サーバ
102
ブラウザ
108

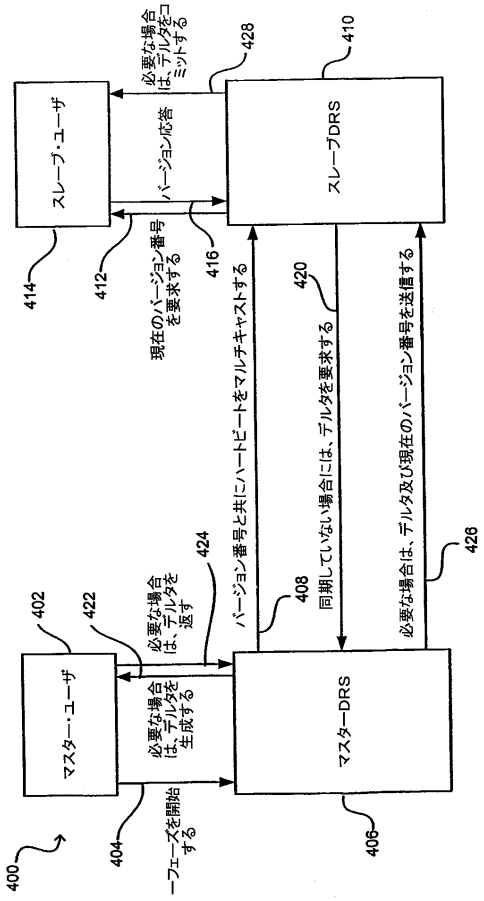
【図２】


210
コンテナ層
スレーブ・ユーザ
スレーブDRS
アップコール
管理クライアント
212
204
ダウンコール
208
アップコール
マスター・ユーザ
マスターDRS
200
202
ダウンコール
206

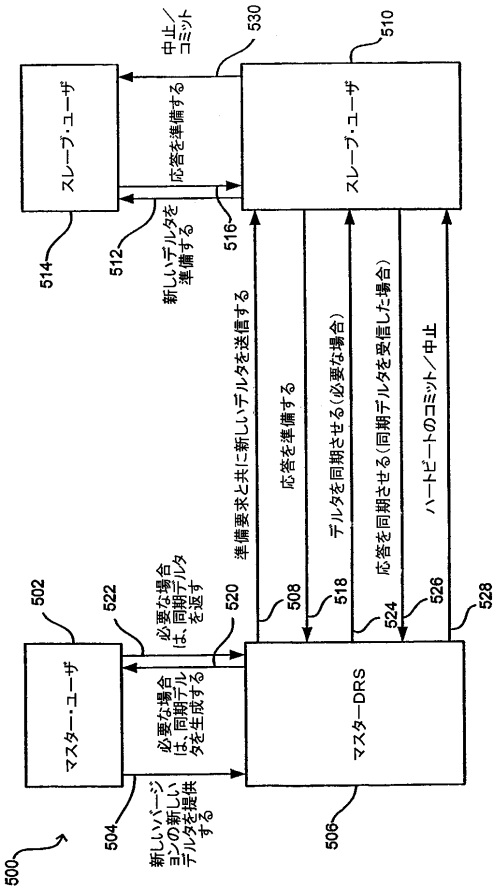
【図３】


300
302
304
ドメイン
管理対象サーバ
308
管理対象サーバ
310
306
ブラウザ
310

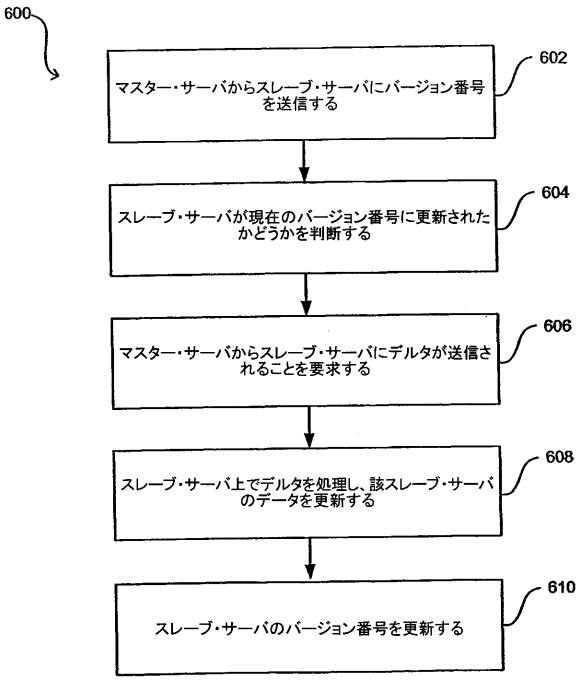
【図４】


必要な場合は、デルタをコミットする
428
410
スレーブ・ユーザ
スレーブDRS
バージョン応答
414
412
現在のバージョン番号を要求する
416
420
バージョン番号と共にハートビートをマルチキャストする
同期していない場合には、デルタを要求する
必要な場合は、デルタ及び現在のバージョン番号を送信する
426
424
408
402
422
必要な場合は、デルタを返す
マスター・ユーザ
必要な場合は、デルタを生成する
マスターDRS
400
404
フェーズを開始する
406

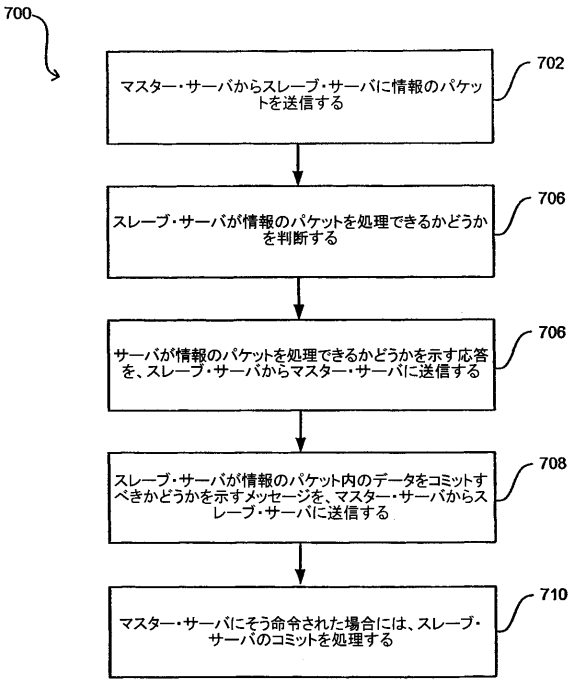
【図５】


スレーブ・ユーザ
中止／コミット
530
応答を準備する
510
スレーブ・ユーザ
514
512
新しいデルタを準備する
516
準備要求と共に新しいデルタを送信する
応答を準備する
デルタを同期させる（必要な場合）
応答を同期させる（同期デルタを受信した場合）
ハートビートのコミット／中止
502
522
必要な場合は、同期デルタを返す
520
508
518
524
526
528
マスター・ユーザ
必要な場合は、同期デルタを生成する
マスターDRS
504
新しいバージョンの新しいデルタを提供する
506
500

【図６】


600
マスター・サーバからスレーブ・サーバにバージョン番号を送信する
602
スレーブ・サーバが現在のバージョン番号に更新されたかどうかを判断する
604
マスター・サーバからスレーブ・サーバにデルタが送信されることを要求する
606
スレーブ・サーバ上でデルタを処理し、該スレーブ・サーバのデータを更新する
608
スレーブ・サーバのバージョン番号を更新する
610

【図７】


700
マスター・サーバからスレーブ・サーバに情報のパケットを送信する
702
スレーブ・サーバが情報のパケットを処理できるかどうかを判断する
706
サーバが情報のパケットを処理できるかどうかを示す応答を、スレーブ・サーバからマスター・サーバに送信する
706
スレーブ・サーバが情報のパケット内のデータをコミットすべきかどうかを示すメッセージを、マスター・サーバからスレーブ・サーバに送信する
708
マスター・サーバにそう命令された場合には、スレーブ・サーバのコミットを処理する
710

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/22366

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : G06F 15/16; G06F 7/00, 17/30
US CL : 709/203; 707/2, 8, 10, 101
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 709/203, 208, 209, 242, 246

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 6,088,694 A (BURNS et al.) 11 JULY 2000 (11.07.2000), col. 4 line 30 to col. 6 line 43.	1-93
A	US 5,920,867 A (VAN HUBEN et al.) 06 JULY 1999 (06.07.1999), col 6 line 50 to col. 8 line 3.	1-93
A, P	US 6,263,372 B1 (HOGAN et al.) 17 JULY 2001, (17.07.2001), col 3 line 23 to col. 9 line 45.	1-93

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
18 November 2002 (18.11.2002)

Date of mailing of the international search report
28 FEB 2003

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703)305-3230

Authorized officer
Ayaz R Sheikh

Telephone No. 703 305 3900

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

PCT/US02/22366

Continuation of B. FIELDS SEARCHED Item 3:
West, Derwent, EPO, JPO
master, slave, client, server, replicating, duplicating, copying, updating, changing, modifying, delta, phase, version

フロントページの続き

(31)優先権主張番号 09/975,587
(32)優先日 平成13年10月11日(2001.10.11)
(33)優先権主張国 米国(US)

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,SK,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EC,EE,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NO,NZ,OM,PH,PL,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TN,TR,TT,TZ,UA,UG,UZ,VN,YU,ZA,ZM,ZW

(74)代理人 100074228
弁理士 今城 俊夫
(74)代理人 100086771
弁理士 西島 孝喜
(72)発明者 ジェイコブス ディーン バーナード
アメリカ合衆国 カリフォルニア州 ９４７０７ バークリー マデラ ストリート １７４７
(72)発明者 ラマー リート
アメリカ合衆国 カリフォルニア州 ９４１３３ サン フランシスコ グリーン ストリート ４１１ ＃２エイ
(72)発明者 スリニヴァサン アナンサム バーラ
アメリカ合衆国 カリフォルニア州 ９４１３１ サン フランシスコ サンチェ ストリート １６１０

Ｆターム(参考) 5B082 GA05 HA02 HA03