

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4267094号
(P4267094)

(45) 発行日 平成21年5月27日(2009.5.27)

(24) 登録日 平成21年2月27日(2009.2.27)

(51) Int. Cl. F I
G06F 12/00 (2006.01) G O 6 F 12/00 5 3 1 D
G06F 11/20 (2006.01) G O 6 F 12/00 5 3 3 F
 G O 6 F 11/20 3 1 0 C

請求項の数 6 (全 24 頁)

(21) 出願番号	特願平10-227291	(73) 特許権者	000155469
(22) 出願日	平成10年8月11日(1998.8.11)		株式会社野村総合研究所
(65) 公開番号	特開2000-57030(P2000-57030A)		東京都千代田区丸の内一丁目6番5号
(43) 公開日	平成12年2月25日(2000.2.25)	(74) 代理人	100075812
審査請求日	平成13年11月12日(2001.11.12)		弁理士 吉武 賢次
審判番号	不服2006-14049(P2006-14049/J1)	(74) 代理人	100091982
審判請求日	平成18年7月3日(2006.7.3)		弁理士 永井 浩之
		(74) 代理人	100096895
			弁理士 岡田 淳平
		(74) 代理人	100117787
			弁理士 勝沼 宏仁
		(72) 発明者	柿 木 彰
			神奈川県横浜市保土ヶ谷区神戸町134番地 株式会社野村総合研究所内

最終頁に続く

(54) 【発明の名称】 2重更新を行うデータベースを有するクライアントサーバーシステム

(57) 【特許請求の範囲】

【請求項1】

少なくとも1つのクライアントと、前記クライアントの接続サーバーを切り替えるゲートウェイと、同一の構成を有する正系サーバーと副系サーバーと、前記正系サーバーと副系サーバーとによってそれぞれ参照および更新される一対のデータベースと、を有するクライアントサーバーシステムにおいて、

前記正系サーバーと副系サーバーはともに、前記クライアントのメッセージに対してオンライン処理を行うオンラインアプリケーション手段と、データベースの参照と更新を行うデータベースシステムと、自系のデータベースに対するデータベース変更メッセージを一時的に蓄積して他系に送出するメッセージキュー送出手段と、他系のデータベースに対するデータベース変更メッセージを自系に受け入れて一時的に蓄積するメッセージキュー受入手段と、前記オンラインアプリケーション手段がデータベース変更メッセージを受けてデータベースに対して行う処理と同一の処理を行うように構成され、前記メッセージキュー受入手段から他系のデータベースのデータベース変更メッセージを入力し、前記データベースシステムを介して自系のデータベースを更新するメッセージキュー起動アプリケーション手段とを有し、

通常の処理においては前記正系サーバーは、そのオンラインアプリケーション手段によってオンライン処理を行い、必要に応じて正系のデータベースシステムを介して正系のデータベースを変更するとともに、正系のデータベースに対するデータベース変更メッセージを正系のメッセージキュー送出手段に送り、

前記副系サーバーは、副系のメッセージキュー受入手段が正系のメッセージキュー送出手段から正系のデータベースに対するデータベース変更メッセージを逐次受け入れ、副系のメッセージキュー起動アプリケーション手段が副系のメッセージキュー受入手段から正系のデータベースに対するデータベース変更メッセージを入力し、副系のデータベースシステムを介して副系のデータベースを更新し、

前記正系サーバーにファイルの破損が発生したときは、前記ゲートウェイの切替えによって処理要求のあった前記クライアントを前記副系サーバーに接続し、副系のオンラインアプリケーション手段によってオンライン処理を行い、必要に応じて副系のデータベースシステムを介して副系のデータベースを変更するとともに、副系のデータベースに対するデータベース変更メッセージを逐次副系のメッセージキュー送出手段に送り、

10

前記正系サーバーは機能が回復した後に、正系のメッセージキュー受入手段が副系のメッセージキュー送出手段から副系処理中の副系のデータベースに対するデータベース変更メッセージを受け入れ、正系のメッセージキュー起動アプリケーション手段が正系のメッセージキュー受入手段から副系のデータベースに対するデータベース変更メッセージを入力し、正系のデータベースシステムを介して正系のデータベースを更新することを特徴とする2重更新を行うデータベースを有するクライアントサーバーシステム。

【請求項2】

前記正系および副系のメッセージキュー起動アプリケーション手段は、起動用アプリケーション制御手段を有し、この起動用アプリケーション手段の起動メッセージによって起動し、前記メッセージキュー受入手段から他系のデータベースに対するデータベース変更メッセージを入力して前記データベースシステムを介して自系のデータベースを更新するように構成されていることを特徴とする請求項1に記載の2重更新を行うデータベースを有するクライアントサーバーシステム。

20

【請求項3】

前記副系サーバーは、副系のメッセージキュー受入手段と平行して正系のデータベースに対するデータベース変更メッセージを入力する予備メッセージキュー受入手段を有することを特徴とする請求項1または2に記載の2重更新を行うデータベースを有するクライアントサーバーシステム。

【請求項4】

前記正系サーバーあるいは副系サーバーと同一のハードウェア構成を有し、前記正系サーバーがサーバー全体の機能停止を伴うハードウェア障害またはプロセスダウンを生じたときに、前記正系サーバーで起動されていたプログラムを起動して前記正系サーバーが行っていた処理を継続して行う待機系サーバーを有することを特徴とする請求項1ないし3のいずれかに記載の2重更新を行うデータベースを有するクライアントサーバーシステム。

30

【請求項5】

前記待機系サーバーは、起動後に正系のデータベースの破損によって処理不能であるときは、前記正系のデータベースに対するデータベース変更メッセージを副系のメッセージキュー受入手段に送出し、

前記副系サーバーは、そのメッセージキュー起動アプリケーション手段が副系のメッセージキュー受入手段から前記正系のデータベースに対するデータベース変更メッセージを入力し、副系のデータベースシステムを介して副系のデータベースを更新した後に、副系のオンラインアプリケーション手段によってオンライン処理を行うことを特徴とする請求項4に記載の2重更新を行うデータベースを有するクライアントサーバーシステム。

40

【請求項6】

前記正系サーバーがオンライン処理した所定期間中の正系のデータベースのデータを集約してデータベースファイルの形式で一時的に記憶する中間ボリューム記憶装置と、

前記副系サーバーがオンライン処理した所定期間中の副系のデータベースに対するデータベース変更メッセージを一時的に記憶しておく副系メッセージキュー手段とを有し、

前記正系サーバーと前記副系サーバーとが交互に切り替わって切れ目なくオンライン処

50

理を行い、切替えによって前記正系サーバーがオンライン処理を開始すると、副系サーバーにおいて、前記中間ボリューム記憶装置に記憶された切替え前の処理期間の開始時点のデータを集約したデータベースファイルを副系のデータベースに上書きし、続いて前記副系メッセージキュー手段に記憶された切替え前の処理期間のデータベース変更メッセージによって副系のデータベースを更新し、次にオンライン処理に伴って生じる正系のデータベースに対するデータベース変更メッセージを逐次入力して副系のデータベースを最新のデータ状態に更新し、

切替えによって前記副系サーバーがオンライン処理を開始すると、正系サーバーにおいて、切替え前の処理期間中に変更された正系データベースのデータの集約と再登録と前記中間ボリューム記憶装置への複写を行うバッチ処理を行い、前記バッチ処理終了後は、オンライン処理に伴って生じる副系のデータベースに対するデータベース変更メッセージを逐次入力して正系のデータベースを最新のデータ状態に更新することにより、24時間連続稼動するように構成したことを特徴とする請求項1ないし5のいずれかに記載の2重更新を行うデータベースを有するクライアントサーバーシステム。

10

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、データベースマネジメントシステム(DBMS)を有するクライアントサーバーシステムに係り、特にハードウェアやデータベースシステムの障害に対して短時間でシステムを復元する障害回復処理を行うクライアントサーバーシステムに関する。

20

【0002】

また、本発明は、24時間連続稼動することができるクライアントサーバーシステムに関する。

【0003】

なお、本発明は、クライアントのリクエストによりサーバーがDBMSを介してデータベースの参照・更新を行うものであればよく、オンライントランザクション処理システムに限られないが、以下の説明ではオンライントランザクション処理システムとして使用した場合にその機能が理解しやすいことが多いため、必要に応じてオンライントランザクション処理システムを例に説明する。

【0004】

【従来の技術】

一般に、オンライントランザクション処理(OLTP)を行うクライアントサーバーシステムは、サーバーに複数のクライアントがオンライン接続し、クライアントからサーバーに処理のためのメッセージを送り、そのメッセージにしたがってサーバーで一連のデータベースアクセス等の処理を行い、その処理結果を即座にクライアントに送り返す処理を行う。

30

【0005】

ここで、オンライントランザクション処理は、クライアントからサーバーに送る1回のメッセージに対する処理を指し、1回のメッセージに対する処理は通常複数のデータベース操作を含み、1つの論理処理単位を形成する。たとえば、データベース内のあるテーブルの値を変更する処理の場合、そのテーブルに関連する他のテーブルの値も変更しなければならない場合には、必要なデータの変更全体が1つトランザクション処理を形成する。このように1つのトランザクション処理が、複数のデータテーブルの値の変更に関係することが多いので、データの値の整合性を確保するため、トランザクション処理は全体として完全に行われるか全く行われないうのいずれかでなければならない。

40

【0006】

また、オンライントランザクション処理を行うクライアントサーバーシステムの場合は、複数のクライアントから不定期にメッセージを受けて上述したトランザクション処理を行う。このため、複数のクライアントから同時に処理要求があった場合には、複数のトランザクション処理を同時に実行しなければならない。このような複雑な処理を行うオンライ

50

ントランザクション処理システムは、多数のトランザクション処理を高速、かつ、データ間の矛盾が生じないように高い信頼性の下で処理することが求められる。

【 0 0 0 7 】

上述した処理に対する高速・高信頼性の要求により、従前はオンライントランザクション処理システムは汎用機からなることが主であった。

【 0 0 0 8 】

汎用機によるオンライントランザクション処理システムでは、使用環境を特化し、種々の処理状態を想定して繰り返しテストすることにより、ハードウェアとソフトウェアの双方においてエラーの発生を最小限にとどめるようにしていた。このため、汎用機によるオンライントランザクション処理システムでは、システム障害が発生する確率がそもそも低い10
ために障害回復処理が必要とされることが少なかった。万一システム障害が発生した場合には、従来の汎用機システムでは予め用意された手順によってシステム回復を図っていたが、その回復処理は通常比較的長い時間を要していた。

【 0 0 0 9 】

これに対して、最近ではネットワークコンピューティングを制御するOS（オペレーティングシステム）によってもオンライントランザクション処理システムを構築できるようになった。ネットワークコンピューティングによるオンライントランザクション処理システムは、汎用機に比べて比較的小型のコンピュータを複数接続してネットワークを構築し、OSの管理の下でトランザクション処理を行うものである。

【 0 0 1 0 】

上記OSの下で構築したオンライントランザクション処理システムにおいても、汎用機同様の処理の高速性と高信頼性が要求される。しかし、ネットワークコンピューティングの処理を管理するOSは、OSとしての汎用性や多機能性を追求するために、トランザクション処理における特定の使用条件における繰り返しテストを行うことはできない。このため、通常のOSには、一般的な障害回復処理の手段が機能として用意されている。一般的な障害回復処理の手段の代表的なものは一部のハードウェアの2重化があり、たとえば、ハードディスクの2重化、SCSIバスの2重化などである。ハードウェアの2重化は基本的なハードウェア障害対策といえる。

【 0 0 1 1 】

これに対して、AT&Tベル研究所の開発によるOSである「UNIX」（UNIXは商品名）はさらに進んでハードウェア（コンピュータ）全体の2重化を想定し、ハードウェア切替用のハイアベイラビリティ機構（HA機構）を用意している。

【 0 0 1 2 】

このHA機構は、サーバーに正系ノード（ノードはクライアントサーバーシステムを構成する論理的単位をいうものとする）と待機系ノードを用意している場合に、正系のノードでLANボード障害や、たとえばマザーボードやCPU障害のようなサーバー本体の全面障害や、データベースシステムなどのプロセスダウン（プロセスダウンとはシステムとしての機能を失うような全体的なシステムダウンをいうものとする）が発生したときに、待機系ノードは障害発生時に正系ノードで処理を行っていたプログラムを起動し、自動的に待機系ノードによって正系ノードの処理を引き継ぐことによってシステムを回復させる機能である。なお、このHA機構による正系と待機系の切り替えは瞬時のうちに行うことができる。

【 0 0 1 3 】

上記HA機構によれば、UNIXによる2重化ハードウェアを有するトランザクション処理システムにおいて、上述したようなハードウェア障害、またはプログラムの異常な処理に伴うシステムのシャットダウンなどに対して、瞬時にノードを切り替え、素早い障害回復処理を行うことができる。

【 0 0 1 4 】

この他に、以上のUNIXのHA機構が、システムのノード自体のハードウェア障害を回復するのに対して通信ネットワークの障害に対してシステムを回復するクライアントサー

10

20

30

40

50

バーシステム（特開平7-302237号）も提案されている。これもハードウェア面におけるシステムの信頼性向上のための提案の一つと言える。

【0015】

この特開平7-302237号が開示するシステムは、クライアントとサーバーを接続するLANを2重化し、通信経路を選択することができるプロセスをそれぞれクライアントとサーバーに置き、一つの通信経路（LAN）に障害があった場合に、他の通信経路（LAN）によって通信を行えるようにしたものである。

【0016】

このシステムによれば、一方の通信ネットワークに障害が発生した場合にも、残る健全な通信ネットワークを選択することにより通信処理を行うことができる。

10

【0017】

【発明が解決しようとする課題】

しかし、従来のクライアントサーバーシステムでは、ハードウェア自体の障害やソフトウェアプログラムのプロセスダウンに伴うハードウェア障害に対しては上述したようにハードウェアの切替えによってシステムの回復を図ることができたが、データベースシステム関連のファイルが破損しているなどのソフトウェアの深刻な障害が生じた場合には、障害回復処理を行うことができなかった。

【0018】

たとえば、上述したUNIXのHA機構は、正系ノードでOLTPのプロセスダウンが発生したときに、正系ノードで処理を行っていたデータベースシステムを待機系ノードで起動することにより、待機系ノードによってシステム回復を図るようにしている。しかし、データベースシステムがもともとプログラムエラー（バグ）を有している場合や、プロセスダウンした時にデータベースの関連ファイルが破損してしまった場合には、これらのデータ検索等を行うプログラム（データベースシステム）を起動した待機系ノードでも、結局データベースシステムを使用することができず、システムを回復することができなかった。

20

【0019】

ところで、上述したようなプログラムエラーやプロセスダウンなどのソフトウェア障害は、通常多数のトランザクション処理を同時に実行している場合に、処理のタイミングによって発生することが多いことが経験上知られている。この場合に、同時実行数が少ないなどトランザクション処理の条件が緩やかになれば、プログラムエラーが発生せずに同一のトランザクション処理をすることが多いことも知られている。

30

【0020】

そこで、本発明は、「常時」最新の状態に更新されるデータベースとそのデータベースシステム（データベースのデータを参照、更新するプログラム）を2重に有していれば、一方のデータベースシステムが破損した場合にも、他方のデータベースとデータベースシステムとによってシステムを回復し継続して処理を行うことができる可能性に着目し、2重のデータベースを常時更新するシステムを提供しようとするものである。

【0021】

ただし、オンライン処理自体は、通常一つのデータベースに対して参照や更新を行うものであるため、一対（2重）のデータベースを常時更新するためには、オンライン処理を実際に行っているシステム以外のシステムが、オンライン処理を行っているシステムの処理結果を常時追跡し、かつ、障害発生時のシステム切替えに伴うデータベース間のデータの矛盾が生じないようにシステム構成する必要がある。このようにオンライン処理システムにおいて2重化されたデータベースをほぼ同時に更新するシステムは従来存在していなかった。

40

【0022】

すなわち、本発明が解決しようとする課題は、ソフトウェア自体の破損、特にデータベースシステムの破損が生じた場合にも、極めて短い時間の間にソフトウェア障害発生前の状態にシステムを回復することができるクライアントサーバーシステムを提供することにあ

50

る。

【 0 0 2 3 】

なお、上述したように、ここで、「システムの回復」というのは、システムとして作動が可能になったことのみならず、障害発生時に完了していなかった処理についてユーザーが受け取ったメッセージとシステムのデータベースの値が互い整合した状態に回復し、また、障害発生時の処理と処理再開後の処理でデータ上矛盾が生じないようにするシステム回復も含むものとする。

【 0 0 2 4 】

さらに、本発明が解決しようとする他の課題として、24時間稼働できるオンライン処理システムを提供することがある。

10

【 0 0 2 5 】

24時間稼働できるオンライン処理システムは、ハードウェアを2重化し、正系システムと副系システムのデータの内容を整合させて、極めて短い時間の間に正系システムと副系システムを切り替えることによって実現されるが、正副両系のデータ内容を整合させること、および極めて短い時間の間に正系システムと副系システムを切り替えることは、技術的には上述した障害回復処理を行うクライアントサーバーシステムと共通である。

【 0 0 2 6 】

ただし、24時間稼働できるためには、上述したような正系・副系システムの迅速な切替えのほかに、2重化したデータベースのそれぞれにおいてデータの整理・再登録の処理が必要であり、このための解決手段が必要である。本発明による24時間稼働できるオンライン処理システムは、正系システムと副系システムを切れ目なく切り替えることができ、かつ、その間に各データベースでデータの整理・再登録をすることができるようにしたものである。

20

【 0 0 2 7 】

すなわち、本発明が解決しようとする他の課題は、障害回復処理を行うクライアントサーバーシステムの延長として24時間稼働することができるオンライン処理システムを提供することにある。

【 0 0 2 8 】

【課題を解決するための手段】

本願請求項1に係る2重更新を行うデータベースを有するクライアントサーバーシステムは、

30

少なくとも1つのクライアントと、前記クライアントの接続サーバーを切り替えるゲートウェイと、同一の構成を有する正系サーバーと副系サーバーと、前記正系サーバーと副系サーバーとによってそれぞれ参照および更新される一対のデータベースと、を有するクライアントサーバーシステムにおいて、

前記正系サーバーと副系サーバーはともに、前記クライアントのメッセージに対してオンライン処理を行うオンラインアプリケーション手段と、データベースの参照と更新を行うデータベースシステムと、自系のデータベースに対するデータベース変更メッセージを一時的に蓄積して他系に送出するメッセージキュー送出手段と、他系のデータベースに対するデータベース変更メッセージを自系に受け入れて一時的に蓄積するメッセージキュー受入手段と、前記オンラインアプリケーション手段がデータベース変更メッセージを受けてデータベースに対して行う処理と同一の処理を行うように構成され、前記メッセージキュー受入手段から他系のデータベースのデータベース変更メッセージを入力し、前記データベースシステムを介して自系のデータベースを更新するメッセージキュー起動アプリケーション手段とを有し、

40

通常の処理においては前記正系サーバーは、そのオンラインアプリケーション手段によってオンライン処理を行い、必要に応じて正系のデータベースシステムを介して正系のデータベースを変更するとともに、正系のデータベースに対するデータベース変更メッセージを正系のメッセージキュー送出手段に送り、

前記副系サーバーは、副系のメッセージキュー受入手段が正系のメッセージキュー送出

50

手段から正系のデータベースに対するデータベース変更メッセージを逐次受け入れ、副系のメッセージキュー起動アプリケーション手段が副系のメッセージキュー受入手段から正系のデータベースに対するデータベース変更メッセージを入力し、副系のデータベースシステムを介して副系のデータベースを更新し、

前記正系サーバーにファイルの破損が発生したときは、前記ゲートウェイの切替えによって処理要求のあった前記クライアントを前記副系サーバーに接続し、副系のオンラインアプリケーション手段によってオンライン処理を行い、必要に応じて副系のデータベースシステムを介して副系のデータベースを変更するとともに、副系のデータベースに対するデータベース変更メッセージを逐次副系のメッセージキュー送出手段に送り、

前記正系サーバーは機能が回復した後に、正系のメッセージキュー受入手段が副系のメッセージキュー送出手段から副系処理中の副系のデータベースに対するデータベース変更メッセージを受け入れ、正系のメッセージキュー起動アプリケーション手段が正系のメッセージキュー受入手段から副系のデータベースに対するデータベース変更メッセージを入力し、正系のデータベースシステムを介して正系のデータベースを更新することを特徴とするものである。

【0029】

本願請求項2に係る2重更新を行うデータベースを有するクライアントサーバーシステムは、請求項1のクライアントサーバーシステムにおいて、

前記正系および副系のメッセージキュー起動アプリケーション手段は、起動用アプリケーション制御手段を有し、この起動用アプリケーション手段の起動メッセージによって起動し、前記メッセージキュー受入手段から他系のデータベースに対するデータベース変更メッセージを入力して前記データベースシステムを介して自系のデータベースを更新するように構成されていることを特徴とするものである。

【0030】

本願請求項3に係る2重更新を行うデータベースを有するクライアントサーバーシステムは、請求項1または2に記載のクライアントサーバーシステムにおいて、

前記副系サーバーは、副系のメッセージキュー受入手段と平行して正系のデータベースに対するデータベース変更メッセージを入力する予備メッセージキュー受入手段を有することを特徴とするものである。

【0031】

本願請求項4に係る2重更新を行うデータベースを有するクライアントサーバーシステムは、請求項1ないし3のいずれかのクライアントサーバーシステムにおいて、

前記正系サーバーあるいは副系サーバーと同一のハードウェア構成を有し、前記正系サーバーがサーバー全体の機能停止を伴うハードウェア障害またはプロセスダウンを生じたときに、前記正系サーバーで起動されていたプログラムを起動して前記正系サーバーが行っていた処理を継続して行う待機系サーバーを有することを特徴とするものである。

【0032】

本願請求項5に係る2重更新を行うデータベースを有するクライアントサーバーシステムは、請求項4のクライアントサーバーシステムにおいて、

前記待機系サーバーは、起動後に正系のデータベースの破損によって処理不能であるときは、前記正系のデータベースに対するデータベース変更メッセージを副系のメッセージキュー受入手段に送出し、

前記副系サーバーは、そのメッセージキュー起動アプリケーション手段が副系のメッセージキュー受入手段から前記正系のデータベースに対するデータベース変更メッセージを入力し、副系のデータベースシステムを介して副系のデータベースを更新した後に、副系のオンラインアプリケーション手段によってオンライン処理を行うことを特徴とするものである。

【0033】

本願請求項6に係る2重更新を行うデータベースを有するクライアントサーバーシステムは、請求項1ないし請求項5のいずれかのクライアントサーバーシステムにおいて、

前記正系サーバーがオンライン処理した所定期間中の正系のデータベースのデータを集約してデータベースファイルの形式で一時的に記憶する中間ボリューム記憶装置と、前記副系サーバーがオンライン処理した所定期間中の副系のデータベースに対するデータベース変更メッセージを一時的に記憶しておく副系メッセージキュー手段とを有し、前記正系サーバーと前記副系サーバーとが交互に切り替わって切れ目なくオンライン処理を行い、切替えによって前記正系サーバーがオンライン処理を開始すると、副系サーバーにおいて、前記中間ボリューム記憶装置に記憶された切替え前の処理期間の開始時点のデータを集約したデータベースファイルを副系のデータベースに上書きし、続いて前記副系メッセージキュー手段に記憶された切替え前の処理期間のデータベース変更メッセージによって副系のデータベースを更新し、次にオンライン処理に伴って生じる正系のデータベースに対するデータベース変更メッセージを逐次入力して副系のデータベースを最新のデータ状態に更新し、

10

切替えによって前記副系サーバーがオンライン処理を開始すると、正系サーバーにおいて、切替え前の処理期間中に変更された正系データベースのデータの集約と再登録と前記中間ボリューム記憶装置への複写を行うバッチ処理を行い、前記バッチ処理終了後は、オンライン処理に伴って生じる副系のデータベースに対するデータベース変更メッセージを逐次入力して正系のデータベースを最新のデータ状態に更新することにより、24時間連続稼動するように構成したことを特徴とするものである。

【0034】

【発明の実施の形態】

20

次に本発明の実施の形態について説明する。

図1ないし図4に本発明による「2重更新を行うデータベースを有するクライアントサーバーシステム」の一実施形態の構成と、通常の処理(図1)、正系システムにハードウェア障害が発生したときの処理(図2)、サーバー全体の機能停止を伴わないソフトウェア障害が発生したときの処理(図3)、および、サーバー全体の機能停止を伴うプロセスダウンが発生しかつデータベース関連ファイルが破損した場合の処理(図4)の流れを示す。

【0035】

図1は、本発明の一実施形態による「2重更新を行うデータベースを有するクライアントサーバーシステム」の構成とその通常処理の流れを示している。

30

【0036】

図1において、2重更新を行うデータベースを有するクライアントサーバーシステム1は、クライアント2a, 2b, ..., 2nと、ゲートウェイ3と、正系サーバー4aと、副系サーバー4bと、正系サーバー4aによって参照・更新される正系データベース5aと、副系サーバー4bによって参照・更新される副系データベース5bと、待機系サーバー6とを有している。

【0037】

クライアント2a, 2b, ..., 2nは、ユーザーの入力により、正系サーバー4aまたは副系サーバー4bに接続し、正系サーバー4aまたは副系サーバー4bに対して所定の処理を要求し、その処理の結果をユーザーに表示または出力する端末装置である。

40

【0038】

ゲートウェイ3は、クライアント2a, 2b, ..., 2nのいずれかから接続要求があったときに、システムの状態により接続先サーバー(正系サーバー4aまたは副系サーバー4b)を切り替えて接続する装置である。

【0039】

正系サーバー4aは、通常処理を行っている時にクライアント2a, 2b, ..., 2nからの処理要求(メッセージ)を処理するサーバーである。正系サーバー4aの内部構成については後述する。正系サーバー4aは、クライアント2a, 2b, ..., 2nの要求に応じて、正系データベース5aのデータをクライアント2a, 2b, ..., 2nに返送したり、必要なデータ処理を行った後に処理結果をクライアント2a, 2b, ..., 2nに返送する装置である。正

50

系サーバー 4 a の内部構成については後述する。

【 0 0 4 0 】

正系データベース 5 a は、上記正系サーバー 4 a によって参照・更新されるデータベースである。正系データベース 5 a は、物理的に正系サーバー 4 a から独立した存在でもよく、また、正系サーバー 4 a 付属の記憶装置内に形成されていてもよい。

【 0 0 4 1 】

副系サーバー 4 b は、正系サーバー 4 a が何らかの原因によって機能停止したときに、正系サーバー 4 a の代替として作動し、クライアントサーバーシステム 1 としての機能を回復するサーバーである。

【 0 0 4 2 】

副系サーバー 4 b は正系サーバー 4 a とほぼ同一の構成（内部の処理手段）を有し、この副系サーバー 4 b の内部構成については後述する。

【 0 0 4 3 】

副系データベース 5 b は、上記副系サーバー 4 b によって参照・更新されるデータベースである。副系データベース 5 b は、後述する処理により、非同期であるが継続的に正系データベース 5 a の変更を追従して更新し、かつ、正系サーバー 4 a 障害時には正系データベース 5 a と同一内容によって、回復したシステムのデータベースとして機能する。なお、副系データベース 5 b は副系サーバー 4 b から独立した存在でもよく、また、副系サーバー 4 b 付属のものでもよいことは、正系データベース 5 a と同様である。

【 0 0 4 4 】

待機系サーバー 6 は、正系サーバー 4 a や副系サーバー 4 b と同一のハードウェアを有し、通常はオペレーティングシステムのみが起動した状態でなんら処理を行わないが、正系サーバー 4 a が障害によって機能停止したときに、正系サーバー 4 a で処理を行っていたプログラムをメモリーに読み込んで実行し、障害発生時に正系サーバー 4 a が行っていた処理を継続して行うサーバーである。なお、上記サーバーの機能停止時に予備システムがそのプログラムを起動して処理を継続するシステム構成は、UNIX においては HA 機構と呼ばれ、汎用機分野ではホットスタンバイと呼ばれている機能である。本発明は、UNIX と同様の機能を有する OS に適用可能であるが、理解容易のために UNIX における「HA 機構」を用語として使用する。

【 0 0 4 5 】

次に正系サーバー 4 a と副系サーバー 4 b の内部構成について説明する。

【 0 0 4 6 】

正系サーバー 4 a は、オンラインアプリケーション手段 7 a と、メッセージキュー起動アプリケーション手段 8 a（図において MQ 起動アプリケーション手段と表示する）と、データベースシステム 9 a と、メッセージキュー送出手段 1 0 a（図において MQ 送出手段と表示する）と、メッセージキュー受入手段 1 1 a（図において MQ 受入手段と表示する）と、起動用アプリケーション制御手段 1 2 a とを有している。

【 0 0 4 7 】

副系サーバー 4 b は正系サーバー 4 a とほぼ同様の構成を有している。すなわち、副系サーバー 4 b は、オンラインアプリケーション手段 7 b と、メッセージキュー起動アプリケーション手段 8 b（図において MQ 起動アプリケーション手段と表示する）と、データベースシステム 9 b と、メッセージキュー送出手段 1 0 b（図において MQ 送出手段と表示する）と、メッセージキュー受入手段 1 1 b（図において MQ 受入手段と表示する）と、起動用アプリケーション制御手段 1 2 b とを有している。ただし、正系サーバー 4 a と異なる点として、本実施形態では副系サーバー 4 b は予備メッセージキュー受入手段 1 3（図において予備 MQ 受入手段と表示する）を有している。

【 0 0 4 8 】

オンラインアプリケーション手段 7 a, 7 b は、クライアント 2 a, 2 b, ..., 2 n と接続してクライアント 2 a, 2 b, ..., 2 n のメッセージに対してオンライン処理を行う手段である。オンラインアプリケーション手段 7 a, 7 b はその処理において、クライアント 2 a,

10

20

30

40

50

2 b, ..., 2 nの要求に応じてデータベースシステム 9 a, 9 bを介してデータベース 5 a, 5 bにアクセスし、そのデータをそのままクライアント 2 a, 2 b, ..., 2 nのユーザーに送って示したり、そのデータを使用して所定の処理を行い、データベースシステム 9 a, 9 bのデータ更新行ったりする。オンラインアプリケーション手段 7 a, 7 bは好ましくは、オブジェクト指向プログラミングにおけるオブジェクトであり、メッセージにより所定の処理を行い、メッセージ発信元に処理結果を返信するように構成される。

【 0 0 4 9 】

なお、オンラインアプリケーション手段 7 a, 7 bがデータベース 5 a, 5 bの更新を行うときは、たとえば所定のテーブルのデータを更新するために関連するすべてのテーブルのデータを矛盾なく更新するものとする。なお、オンラインアプリケーション手段 7 a, 7 bは、処理の内容に応じてそれぞれ複数種類存在する。また、正系と副系のオンラインアプリケーション手段 7 a, 7 bは同一の内容を有し、同一の処理を行えるようにする。

10

【 0 0 5 0 】

メッセージキュー起動アプリケーション手段 8 a, 8 bは、メッセージキュー受入手段 1 1 a, 1 1 bに格納された他系のデータベースに対するデータベース変更メッセージにしたがってデータベースシステム 9 a, 9 bを介して自系のデータベース 5 a, 5 bを変更する手段である。

【 0 0 5 1 】

メッセージキュー起動アプリケーション手段 8 a, 8 bは、独自のプログラミングと制御により、単独で上述したようにメッセージキュー受入手段 1 1 a, 1 1 bからデータベース 5 a, 5 bの変更メッセージを入力し、データベース 5 a, 5 bの変更を行うようにしてもよいが、好ましくは図 1 に示すように、起動用アプリケーション制御手段 1 2 a, 1 2 bを有し、オンラインアプリケーション手段 7 a, 7 bと同一の処理を行うようにする。

20

【 0 0 5 2 】

このように構成することにより、メッセージキュー起動アプリケーション手段 8 a, 8 bは実質的にはオンラインアプリケーション手段 7 a, 7 bの複写で簡単に形成でき、これに起動を制御する簡単な起動用アプリケーション制御手段 1 2 a, 1 2 bを付加するだけで上記機能を果たすことができるのである。

【 0 0 5 3 】

すなわち、オンラインアプリケーション手段 7 a, 7 bを複写することにより、メッセージキュー起動アプリケーション手段 8 a, 8 bはデータベース 5 a, 5 bに対してオンラインアプリケーション手段 7 a, 7 bとまったく同一の処理を行うことができる。このように、メッセージキュー起動アプリケーション手段 8 a, 8 bがオンラインアプリケーション手段 7 a, 7 bと同様の処理を行うことができることにより、他系のデータベースの変更部分を自系のデータベースに反映でき、同期（データを一致させること）を取ることができるのである。

30

【 0 0 5 4 】

データベースシステム 9 a, 9 bは、オンラインアプリケーション手段 7 a, 7 bの要求（データベースに対する処理要求メッセージ）を受け、データベースのデータを操作する手段である。データベースシステム 9 a, 9 bは、広い意味における「データベース」に含まれているので、データベース 5 a, 5 bの一部になっている場合も本発明のシステムに含まれる。

40

【 0 0 5 5 】

メッセージキュー送出手段 1 0 a, 1 0 bは、自系のデータベースに対するデータベース変更メッセージを一時的に蓄積して他系に送出する手段である。

【 0 0 5 6 】

メッセージキュー受入手段 1 1 a, 1 1 bは、他系のメッセージキュー送出手段 1 0 a, 1 0 bから上述した他系のデータベースに対するデータベース変更メッセージを受け入れて一時的に蓄積する手段である。

【 0 0 5 7 】

50

起動用アプリケーション制御手段 1 2 a, 1 2 bは、既述したように、メッセージキュー起動アプリケーション手段 8 a, 8 bに起動メッセージを送ってこれらを起動するための手段である。既述したようにメッセージキュー起動アプリケーション手段 8 a, 8 bがオンラインアプリケーション手段 7 a, 7 bの複写からなるときは、起動用アプリケーション制御手段 1 2 a, 1 2 bはオンラインアプリケーション手段 7 a, 7 bに対する起動メッセージと同様な起動メッセージをメッセージキュー起動アプリケーション手段 8 a, 8 bに送ってこれらを起動する。

【 0 0 5 8 】

予備メッセージキュー受入手段 1 3 は、正系データベース 5 aと副系データベース 5 bが何らかの原因によって同時に障害発生した時に備え、正系データベース 5 aに対するデータベース変更メッセージを記録しておく手段である。

10

【 0 0 5 9 】

以上が正系サーバー 4 aと副系サーバー 4 bの内部の構成であったが、以下に通常の処理における諸構成手段の処理について説明する。

【 0 0 6 0 】

図 1 にクライアントサーバーシステム 1 の通常の処理の流れを示す。通常の処理においては、正系サーバー 4 aがクライアント 2 a, 2 b, ..., 2 nとオンライン接続して処理を行う。正系サーバー 4 aでは、オンラインアプリケーション手段 7 aが、クライアント 2 a, 2 b, ..., 2 nの要求に応じてオンライン処理を行い、必要に応じて正系のデータベースシステム 9 aを介して正系データベース 5 aを変更するとともに、正系データベース 5 aに対するデータベース変更メッセージを逐次正系のメッセージキュー送出手段 1 0 aに送る。

20

【 0 0 6 1 】

一方正系サーバー 4 aがオンライン処理を行う間、副系サーバー 4 bでは、ディファードオンライン処理で正系データベース 5 aの変更部分を副系データベース 5 bに反映して更新する。

【 0 0 6 2 】

すなわち、副系サーバー 4 bでは、副系のメッセージキュー受入手段 1 1 bが正系のメッセージキュー送出手段 1 0 aから正系データベース 5 aの変更部分を（データベース変更メッセージとして）逐次受け入れ、副系のメッセージキュー起動アプリケーション手段 8 bがその副系のメッセージキュー受入手段 1 1 bから正系データベース 5 aに対するデータベース変更メッセージを入力し、副系のデータベースシステム 9 bを介して副系データベース 5 bを更新する。

30

【 0 0 6 3 】

上記処理により、正系サーバー 4 aでクライアント 2 a, 2 b, ..., 2 nの要求に応じて正系データベース 5 aの変更を行った場合には、そのデータベースの変更が、送出手段 1 0 a（正系）と、受入用メッセージキュー蓄積手段 1 1 b（副系）と、メッセージキュー起動アプリケーション手段 8 b（副系）と、データベースシステム 9 b（副系）とを経て、極めて短い時間の後に副系データベース 5 bに反映される。これにより、副系データベース 5 bを正系データベース 5 aはほぼ同時に同期（データベースの内容を一致させること）され、正系データベース 5 aと副系データベース 5 bの内容はデータベース変更メッセージを送受信するための時間を除き、ほぼ常時最新の状態に一致される。

40

【 0 0 6 4 】

次に、この 2 重更新を行うデータベースを有するクライアントサーバーシステム 1 において、正系サーバー 4 aにハードウェア障害を生じた場合の障害回復処理について図 2 を用いて以下に説明する。

【 0 0 6 5 】

図 2 に正系サーバー 4 aがハードウェア障害を生じた場合の障害回復処理を示す。なお図 2 において、正系サーバー 4 aのハードウェア障害（データベースシステム 9 a等は健全な状態）を表すために正系サーバー 4 aに破線の対角線を付す。

【 0 0 6 6 】

50

正系サーバ 4aがサーバ全体の機能停止を伴うハードウェア障害を生じた場合には、HA機構（「HA機構」はUNIXの場合の呼称であって、他のOSによる場合はUNIXのHA機構に相当する機能）により、待機系サーバ 6がハードウェア障害時に正系サーバ 4aで起動されていたプログラムを起動し、継続処理のために障害発生時のデータファイルを参照する。この処理により待機系サーバ 6には、オンラインアプリケーション手段 7a、メッセージキュー起動アプリケーション手段 8a、データベースシステム 9a、メッセージキュー送出手段 10a、メッセージキュー受入手段 11a、起動用アプリケーション制御手段 12aとそれぞれ同一のオンラインアプリケーション手段 7a'、メッセージキュー起動アプリケーション手段 8a'、データベースシステム 9a'、メッセージキュー送出手段 10a'、メッセージキュー受入手段 11a'、起動用アプリケーション制御手段 12a'が形成される。

10

【0067】

待機系サーバ 6は正系サーバ 4aで行っていた処理を継続して行い、データベースシステム 9a'により、継続して正系データベース 5aに対して参照及び更新を行う。

【0068】

図2に表す場合の正系サーバ 4aのハードウェア障害は、ハードウェア上の障害であってプログラムやデータは健全であるので、上記オンラインアプリケーション手段 7a'、メッセージキュー起動アプリケーション手段 8a'、データベースシステム 9a'、メッセージキュー送出手段 10a'、メッセージキュー受入手段 11a'、起動用アプリケーション制御手段 12a'はそれぞれ健全に機能し、待機系サーバ 6は正系サーバ 4aとまったく同一の処理を行うことができるようになる。

20

【0069】

待機系サーバ 6が正系サーバ 4aと同様の処理を行うサーバとして起動した後は、ゲートウェイ 3は、要求処理のあったクライアント 2a, 2b, ..., 2nを待機系サーバ 6に接続し、待機系サーバ 6によりオンライン処理を継続することができる。

【0070】

上記正系サーバ 4aから待機系サーバ 6への切替えは極めて短い時間内、かつ、自動的にハードウェア障害を検出して行うことができる。このため、正系サーバ 4aのハードウェア障害に対しては、HA機構により短時間の切替処理の経過後にシステムとして障害回復をすることができるのである。

30

【0071】

なお、上記正系サーバ 4aのハードウェア障害発生時の副系データベース 5bのデータ更新は、正系サーバ 4aが待機系サーバ 6に切り替わっただけで副系サーバ 4bにおける処理は何ら変わらない。すなわち、メッセージキュー送出手段 10aに蓄積された正系データベース 5aの変更メッセージは、そのまま待機系サーバ 6のメッセージキュー送出手段 10a'に継承され、システムの障害回復後に当該データベース変更メッセージはメッセージキュー送出手段 10a'からメッセージキュー受入手段 11bと予備メッセージキュー受入手段 13に送られ、メッセージキュー起動アプリケーション手段 8bとデータベースシステム 9bとによって副系データベース 5bに反映され、副系データベース 5bが正系データベース 5aと同一内容になるように更新される。

40

【0072】

以上は、正系サーバ 4aのサーバ機能停止を伴うハードウェア障害に対するシステムの障害回復処理であったが、次に正系サーバ 4aのサーバとして機能は損なわれないが、正系データベース 5aをはじめとするデータベース関連のファイル破損が発生した場合処理について図3を用いて以下に説明する。

【0073】

図3に正系サーバ 4aの正系データベース 5aが破損した場合の障害回復処理の流れを示す。図3において、正系データベース 5aの破損を示すために正系データベース 5aに破線の対角線を付す。

【0074】

50

正系サーバ 4 a の正系データベース 5 a が破損した場合は、正系データベース 5 a のデータの参照及び更新が不可能になるので、副系サーバ 4 b に切り替えてオンライン処理を行い、必要に応じて副系データベース 5 b の参照及び更新を行う。正系データベース 5 a の機能回復後は、副系サーバ 4 b のオンライン処理間に生じた副系データベース 5 b の変更を正系データベース 5 a に反映して、正系データベース 5 a を更新する。

【 0 0 7 5 】

具体的には、正系サーバ 4 a の正系データベース 5 a 等が破損したときは、ゲートウェイ 3 の切り替えによって、処理要求のあったクライアント 2 a , 2 b , ... , 2 n を副系サーバ 4 b に接続する。副系サーバ 4 b は、図 1 で説明したように、副系データベース 5 b が正系データベース 5 a と一致するように常時更新されているので、直ちにあるいは副系データベース 5 b と正系データベース 5 a の同期を確認する極めて短い時間の経過後に副系データベース 5 b を使用してオンライン処理を開始することができる。正系サーバ 4 a から副系サーバ 4 b への切替え自体は極めて短い時間の間に行うことができる。このようにして副系サーバ 4 b の処理を開始することにより、2 重更新を行うデータベースを有するクライアントサーバシステム 1 としては機能が回復し、見かけ上障害回復処理が実現することができるのである。

10

【 0 0 7 6 】

副系サーバ 4 b によるオンライン処理では、オンラインアプリケーション手段 7 b はオンラインアプリケーション手段 7 a と同一の処理を行うことができるので、クライアント 2 a , 2 b , ... , 2 n の要求に応じてデータベースシステム 9 b を介して必要に応じて副系データベース 5 b を参照及び更新する。副系データベース 5 b のデータを変更する場合には、そのデータベース変更メッセージをデータベースシステム 9 b とともに副系のメッセージキュー送出手段 1 0 b にも出力する。

20

【 0 0 7 7 】

上記副系のオンラインアプリケーション手段 7 b は正系のオンラインアプリケーション手段 7 a と同一の処理を行うことができ、かつ、ゲートウェイ 3 による正系副系の切替えは瞬時に行われるので、ユーザーは、正系サーバ 4 a と副系サーバ 4 b の切り替えを意識することなく、オンライン処理をすることができるのである。

【 0 0 7 8 】

副系サーバ 4 b によってオンライン処理を行っている間に、システムエンジニアにより正系サーバ 4 a の障害を回復するようにする。たとえば、破損した正系データベース 5 a を破棄し、障害発生前の健全な状態のデータベースに置き換え、障害発生までのデータの変更を追跡更新するなどの作業を行うことができる。

30

【 0 0 7 9 】

上述したような正系サーバ 4 a の回復作業により正系サーバ 4 a の機能が回復した後は、副系サーバ 4 b の処理中に副系データベース 5 b に生じたデータベースの変更部分を正系データベース 5 a に反映させなければならない。具体的には、データベースシステム 9 a が回復した後に、正系のメッセージキュー受入手段 1 1 a が、副系のメッセージキュー送出手段 1 0 b から副系処理中の副系データベース 5 b のデータベース変更メッセージを入力する。次に、正系のメッセージキュー起動アプリケーション手段 8 a が、起動用アプリケーション制御手段 1 2 a の起動メッセージによって起動し、メッセージキュー受入手段 1 1 a から副系データベース 5 b のデータベース変更メッセージを入力し、正系のデータベースシステム 9 a を介して、正系データベース 5 a を更新する。これにより、正系データベース 5 a は、障害発生中のデータベースに対する変更を反映し、最新の状態である副系データベース 5 b と同期することができる。

40

【 0 0 8 0 】

正系データベース 5 a が更新された後は、再びゲートウェイ 3 の切り替えにより、それ以降に処理要求があったクライアント 2 a , 2 b , ... , 2 n を正系サーバ 4 a に接続し、通常の処理の流れに戻すことができる。

【 0 0 8 1 】

50

以上はサーバー機能停止を伴わないがデータベース関連ファイルが破損した時の障害回復処理であったが、次にサーバー全体の機能停止を伴うソフトウェア障害が生じた場合の障害回復処理について図4を用いて以下に説明する。

【0082】

図4にサーバー全体の機能停止およびデータベース関連ファイルの破損が発生した場合の障害回復処理の流れを示す。なお図4において、サーバー全体の機能停止およびデータベース関連ファイルの破損を表すために、正系サーバー4aと正系データベース5aにそれぞれ破線の対角線を付す。なお、サーバー全体の機能停止を伴うソフトウェア障害としては、ハードウェア障害に伴ってデータベース関連のファイルも破損したような場合や、データベースシステムのプログラム上の不具合(バグ)等により処理中に全面的なファイル破損が生じたような場合などがある。

10

【0083】

正系サーバー4aがソフトウェア障害によってサーバーとしての機能がダウンしたときは、既述したようにHA機構がこれを検知して作動し、待機系サーバー6に障害発生時の正系サーバー4aの処理プログラムを起動させ、待機系サーバー6が自動起動する。HA機構により、オンラインアプリケーション手段7a、メッセージキュー起動アプリケーション手段8a、データベースシステム9a、メッセージキュー送出手段10a、メッセージキュー受入手段11a、起動用アプリケーション制御手段12aは、それぞれ待機系サーバー6にオンラインアプリケーション手段7a'、メッセージキュー起動アプリケーション手段8a'、データベースシステム9a'、メッセージキュー送出手段10a'、メッセージ

20

【0084】

しかし、この場合は起動されたデータベースシステム9a'が参照する正系データベース5aも破損しているので、待機系サーバー6が正系サーバー4aとして機能することができない。

【0085】

そこで、待機系サーバー6からさらに副系サーバー4bに切り替えて、副系サーバー4bによってシステムの回復を図る必要がある。

【0086】

ここで、副系サーバー4bによってシステムを回復するには、その前に待機系サーバー6に滞留している正系データベース5aのデータベース変更メッセージを副系サーバー4bに送り出し、副系データベース5bを更新して障害発生時の正系データベース5aと同一内容としなければならない。この滞留したデータベース変更メッセージの送出と副系データベース5bの更新の処理の流れを図4で一点鎖線の矢印によって示す。

30

【0087】

上記滞留したデータベース変更メッセージを副系サーバー4bに送出して副系データベース5bを更新するには、短時間ではあるがオンライン処理を一時的に中止する。しかる後に、メッセージキュー送出手段10aを起動したメッセージキュー送出手段10a'から副系のメッセージキュー受入手段11bへデータベース変更メッセージを送出する。

【0088】

次に、副系のメッセージキュー起動アプリケーション手段8aは、起動用アプリケーション制御手段12bによって起動し、メッセージキュー受入手段11bから正系データベース5aに対するデータベース変更メッセージ(メッセージキュー送出手段10a'の滞留分)を入力し、副系のデータベースシステム9bを介して副系データベース5bを更新する。この更新により、副系データベース5bは、障害発生直前の正系データベース5aと同一の内容になり、これを使用してオンライン処理が可能な状態になる。

40

【0089】

以上の準備の後、システムはオンライン処理を再開し、ゲートウェイ3の切替えにより、処理要求があったクライアント2a, 2b, ..., 2nを副系サーバー4bに接続する。

【0090】

50

副系サーバー 4bにおいては、オンラインアプリケーション手段 7bがクライアント 2a, 2b, ..., 2nの要求に応じて、必要によりデータベースシステム 9bを介して副系データベース 5bを参照しあるいは更新する。副系データベース 5bは、処理開始時には障害発生直前の正系データベース 5aと同一内容に更新されているので、データに矛盾を生じることなくオンライン処理を再開することができるのである。

【0091】

副系処理中に副系データベース 5bのデータを更新する場合には、オンラインアプリケーション手段 7bはデータベース変更メッセージをデータベースシステム 9bとメッセージキュー送出手段 10bの双方に同時に送るようにする。

【0092】

副系サーバー 4bによって以上のようなオンライン処理を行う間に、システムエンジニア等により、正系サーバー 4aの機能回復を別途図るようにする。正系サーバー 4aの機能が回復した後は、正系のメッセージキュー受入手段 11aが副系のメッセージキュー送出手段 10bから、副系サーバー 4b処理中の副系データベース 5bに対するデータベース変更メッセージを入力する。次に、正系のメッセージキュー起動アプリケーション手段 8aが起動用アプリケーション制御手段 12aの起動メッセージによって起動し、正系のメッセージキュー受入手段 11aから副系処理中の副系データベース 5bに対するデータベース変更メッセージを入力し、データベースシステム 9aを介して正系データベース 5aを更新する。これにより、正系データベース 5aは、最新のデータ内容である副系データベース 5bと同期される。

【0093】

正系データベース 5aと副系データベース 5bの同期をとった後は、正系データベース 5aはオンライン処理が可能な状態になるので、ゲートウェイ 3の切替えによって、通常の処理に戻すことができるのである。

【0094】

従来の汎用機を使用したトランザクション処理システムは、テストを繰り返すことによりシステム停止を伴うソフトウェア障害が発生しないようにしていたが、万一深刻なソフトウェア障害によってシステム機能の全体がダウンした場合には、長時間システムを停止し、全面的な復旧をしなければならなかった。また、従来のUNIXのHA機構を利用した障害回復機能を有するシステムは、単にハードウェアの2重化を図ったものであり、ハードウェア障害時に予備のシステムが起動するが、ハードウェア障害に伴ってデータベース関連ファイルのような重要なソフトウェアに障害が発生した場合は、起動した予備システムによっても結局システムの機能を回復することができず、汎用機の場合と同様にシステムの回復まで長い時間がかかることになった。

【0095】

これら従来のシステムに対し、本発明による「2重更新を行うデータベースを有するクライアントサーバーシステム 1」によれば、ハードウェアのみの障害(図2の場合)、データベース破損等のソフトウェア障害(図3の場合)、ハードウェア障害およびデータベースは損等のソフトウェア障害(図4の場合)のいずれに対しても、迅速にシステムを回復することができ、これによって安価なシステムによって高い信頼性を有するトランザクション処理システムを実現することができるのである。

【0096】

すなわち、本発明の「2重更新を行うデータベースを有するクライアントサーバーシステム 1」によれば、UNIXのような汎用的なオペレーションシステムによって、汎用機に比較して小型・廉価のコンピュータを組み合わせることでオンライン処理システムを構築でき、かつ、上述したように2重化したデータベースの双方をほぼ常時最新の状態に維持することができることにより、片方のサーバーにハードウェア障害やソフトウェア障害が発生した時に他方のサーバーに切り替えて継続して処理でき、高い信頼性のオンライン処理システムを得ることができるのである。

【0097】

10

20

30

40

50

なお、極めて稀なことではあるが、正系サーバー 4a と副系サーバー 4b が同時にもっとも深刻なソフトウェア障害を伴うハードウェア障害となった場合でも、データベース変更メッセージを貯留した予備メッセージキュー受入手段 13 とゲートウェイ 3 の内部の電文トレースにより、システムを回復することができる。この場合は、上述した正系・副系の切替えによるシステム回復より長い時間を必要とするが、正系サーバー 4a と副系サーバー 4b が同時に機能しないケースは非常に稀であるので回復に多少時間がかかる点はやむを得ないものとする。

【0098】

以上は障害回復処理を迅速に行うクライアントサーバーシステムについて説明したが、次に 24 時間連続稼動するクライアントサーバーシステムをについて以下に説明する。

10

【0099】

一般的に 24 時間連続稼動するオンライン処理システムは、同一システムを長期間使用した場合のオペレーティングシステムの不安定な動作、データベースファイルのフラグメンテーション、インデックスツリーの変形による処理性能の劣化等の問題を解決しつつ、24 時間連続稼動することができるようにしなければならない。

【0100】

図 1 ~ 図 4 において説明した 2 重更新データベースを使用するクライアントサーバーシステム 1 は、2 重化したシステム（正系サーバーと副系サーバー）を有し、一方のシステムによってオンライン処理でき、必要なときにオンライン処理を行うシステムを切り替えることができる。この図 1 ~ 図 4 のクライアントサーバーシステムは、一方のシステムが稼動中に他方のシステムについて種々の保守が可能である点で 24 時間連続稼動するオンライン処理システムとしての基本的な機能を有している、ということができる。

20

【0101】

ただし、図 1 ~ 図 4 のクライアントサーバーシステムは、一方のシステムのデータベースの変更を逐一他方のシステムのデータベースに反映するようにしているので、「データの洗い換え」の処理を行っていない。以下に説明する 24 時間連続稼動するオンライン処理システムは、すでに説明した「2 重更新データベースを使用するクライアントサーバーシステム」にデータの洗い換えをするための中間ボリューム記憶装置を付加し、データの洗い換えという処理を行うようにしたものである。

【0102】

「データの洗い換え」の処理は、同一のデータに対する複数回の処理を集計し、不要な途中のデータを整理することによってデータ量を圧縮する処理である。

たとえば、銀行預金を管理するオンライン処理システムにおいては、あるユーザーが自分の口座から引出しをするのに、小額の預金を多数回引き出すことがしばしばみられる。この場合、預金を管理するデータベースでは、預金の引出し回数と同じ回数の取引データが記録され、データベースが管理するデータ数が増加する。

30

【0103】

しかし、預金を管理するデータベースとして必要なデータは、最新の残高である場合が多いので、途中の取引経過を整理し、最新の残高データのみを登録しておけばよいことがある。このような場合に、途中の取引経過を集計し、データ量を圧縮して最新の残高のみを登録する処理が「データの洗い換え」の処理である。銀行預金を管理するデータベースに限られず、一般にオンライン処理を行うシステムでも、データベースの無限な膨張を防止するために、一定の期間ごとにデータを整理し、再登録する処理（データの洗い換え処理）を行うことが必要となる。本発明による 24 時間連続稼動するオンライン処理システムでは、上記「データの洗い換え」の処理を正系サーバーと副系サーバーの切替え中に行うようにしている。

40

【0104】

まず、本発明による 24 時間連続稼動するオンライン処理システムの構成について説明する。

【0105】

50

図5, 6に2重更新を行うデータベースを有しオンライン処理を行う本発明による24時間連続稼動するクライアントサーバシステム21(以下簡単に指し示す場合は24時間連続稼動オンライン処理システム21または単にオンライン処理システム21と略称する)の一実施形態の構成と、正系サーバ稼動時の処理の流れ(図5)と、副系サーバ稼動時の処理の流れ(図6)を示す。なお、理解容易のために図5と図6において図1~図4と同一部分については同一符号を付して示し、重複する説明を省略する。

【0106】

図5に示すように、本実施形態による24時間連続稼動オンライン処理システム21は、図1~図4の2重更新データベースを使用するクライアントサーバシステム1とほぼ同一の構成を有している。ただし、本実施形態による24時間連続稼動オンライン処理システム21は、図1~図4に示した2重更新データベースを使用するクライアントサーバシステム1の構成に追加して中間ボリューム記憶装置22と副系メッセージキュー手段23(図5において正系サーバ4a中に副系MQ手段23と示す)とを有している。なお、図5において、図1~図4に示したクライアントサーバシステム1の待機系サーバ6と予備受入用メッセージキュー蓄積手段13を示していないが、これはこれらの諸手段を除かなくてはならないという意味ではなく、単に装置を簡潔に示すためであり、無論これらの諸手段を含んでいてもよい。

【0107】

中間ボリューム記憶装置22は、システム構成上正系データベース5aと副系データベース5bの間に位置し、正系サーバ4aが稼動した所定期間中の正系データベース5aの更新部分を集約してデータベースファイルの形で一時的に記憶し、それを副系サーバ4bに受け渡すための記憶装置である。

【0108】

副系メッセージキュー手段23は、副系サーバ4bが稼動した所定期間中の副系データベース5bへのデータベース変更メッセージを一時的に記憶しておく手段である。本実施形態では、副系メッセージキュー手段23は、副系サーバ4bの内部に設けられている。

【0109】

次に、図7を用いて本実施形態による24時間連続稼動オンライン処理システム21の全体の処理の流れを説明する。

【0110】

図7は、正系サーバ4aと副系サーバ4bの稼動の切替えのタイムチャートを示している。図7において、横軸は時間経過を示し、縦に正系サーバ4aと中間ボリューム記憶装置22と副系サーバ4bを配列して示している。

【0111】

図7の期間0~3はそれぞれ適当に定めた期間であり、各期間0~3の移行ごとに正系サーバ4aまたは副系サーバ4bのいずれかが切り替ってオンライン処理を行う。図7の例では期間0は副系サーバ4b、期間1は正系サーバ4a、期間2は副系サーバ4b、期間3は正系サーバ4aがオンライン処理をしている。サーバ間の切替えはユーザーがシステムの停止を感じることがないほど極めて短い時間のうちに行われる。

【0112】

正系サーバ4aがオンライン処理をしている間(期間1, 3, ...)は、副系サーバ4bは、バッチ的に行うデータの洗い換えと更新と、正系データベース5aの変更に追従して行う更新(この追従更新を便宜上キャッチアップという)を行っている。

【0113】

今、期間2に注目すると、期間2ではオンライン処理が副系サーバ4bに切り替わり、副系サーバ4bがオンライン処理を行い、正系サーバ4aではその前の期間1のオンライン処理中に生じた正系データベース5a(図示せず)のデータ変更を集約し(洗い換え処理をし)これを中間ボリューム記憶装置22に複写するバッチ処理を行う(処理4)。その後、正系サーバ4aは副系サーバ4bのオンライン処理に伴い副系データベー

10

20

30

40

50

ス5b(図示せず)の変更を追随し、キャッチアップする(処理 5)。

【0114】

次に、期間2から期間3に移行すると、オンライン処理が正系サーバー4aに切り替わり、正系サーバー4aがオンライン処理を行う。正系サーバー4aは、期間2中にその正系データベース5a(図示せず)が常に副系データベース5bの更新に追随して更新しているので、期間3の開始時点において最新のデータの状態でオンライン処理を開始することができ、オンライン処理の切替えに障害が生じることがない。

【0115】

一方期間3中、副系サーバー4bは、データの洗い換えと最新状態への更新を行う。オンライン処理サーバーの切替え直後から、副系サーバー4bは、先に期間2で正系サーバー4aが中間ボリューム記憶装置22に複写したデータベースファイルを副系データベース5b(図示せず)に上書きする(処理 1)。この上書きされるデータは、期間1までのデータを洗い換え、すなわち集約して再登録したものであるため、処理 1 が完了した状態では、副系データベース5b(図示せず)は、期間1の終了時点のデータ状態になる。次に、副系データベース5bには、期間2の副系サーバー4bによるオンライン処理のデータ変更部分を追加して更新する(処理 2)。この処理 2 が完了した状態では、副系データベース5b(図示せず)は、期間1の変更データについては洗い換え処理をしていないものの、期間1の終了時点のデータ状態になる。処理 2 が完了後、副系サーバー4bは、正系サーバー4aのオンライン処理に伴い、正系データベース5aに生じたデータ変更部分を逐次的に入力し、副系データベース5bを正系データベース5aに追随して更新する(処理 3)。この処理 3 により、副系データベース5bは、極めて短い時間において正系データベース5aの変更を追随し、ほぼ常時最新のデータ状態に更新される。

【0116】

以上が24時間連続稼動オンライン処理システム21のオンライン処理サーバーの切替え、データベースの2重更新の処理であったが、次にこれらの処理が正系サーバー4aと副系サーバー4b等の構成手段間でどのように処理されるかについて図5と図6を用いて以下に説明する。なお、理解を容易にするために図5と図6においてデータベースの更新処理については図7と同一の符号 1 ~ 5 を付す。

図5は、正系サーバー4aがオンライン処理をしている状態を示している。オンライン処理が正系サーバー4aに切り替わった直後から、中間ボリューム記憶装置22のデータベースファイルが副系データベース5bに上書きされる(処理 1)。処理 1 により、副系データベース5bは、その前の処理期間(注目している期間の前の副系サーバー4bがオンライン処理を行った期間)の当初のデータ状態になる。ただし、データは洗い換え処理を完了した状態に整理されている。処理 1 の完了後は、中間ボリューム記憶装置22と副系データベース5bの接続が切り離される。

【0117】

次に、副系メッセージキュー手段23に蓄積されたデータベース変更メッセージにより、副系データベース5bのデータが更新される(処理 2)。副系メッセージキュー手段23に蓄積されたデータベース変更メッセージは、後に説明するように、切替え前の処理期間(注目している期間の前の副系サーバー4bがオンライン処理を行った期間)に副系データベース5bに対するデータベース変更メッセージを蓄積したものである。処理 2 により、副系データベース5bは、切替え前の処理期間(注目している期間の前の副系サーバー4bがオンライン処理を行った期間)の終了時点のデータ状態になる。

【0118】

次に、副系サーバー4bは、正系サーバー4aのオンライン処理に伴って生じるデータベース変更メッセージを逐次入力して、正系データベース5aの更新を追随して極めて短い時間の経過後副系データベース5bのデータを更新する(処理 3)。処理 3 により、副系データベース5bは、ほぼ常時最新のデータ状態を維持することができる。

【0119】

10

20

30

40

50

上記処理 3 で副系データベース 5 b の更新に使用されるデータベース変更メッセージは、図 5 に示すように、正系の送出用メッセージキュー蓄積手段 1 0 a から副系の受入用メッセージキュー蓄積手段 1 1 b へ送られ、副系のメッセージキュー起動アプリケーション手段 8 b が起動用アプリケーション制御手段 1 2 b によって起動して受入用メッセージキュー蓄積手段 1 1 b から入力し、データベースシステム 9 b を介して副系データベース 5 b を更新するのに使用されるものである。

【 0 1 2 0 】

この間オンライン処理を行っている正系サーバー 4 a では、オンライン処理の切替え後から、オンラインアプリケーション手段 7 a がオンライン処理を行い、データベースシステム 9 a を介して正系データベース 5 a にアクセスし、正系データベース 5 a を変更する場合は、そのデータベース変更メッセージを自系のデータベースシステム 9 a とともに送出用メッセージキュー蓄積手段 1 0 a にも送る。送出用メッセージキュー蓄積手段 1 0 a に送られたこれらのデータベース変更メッセージが上述したように副系データベース 5 b の更新に使用され、これによって副系データベース 5 b が最新のデータ状態に維持されるのである。

10

【 0 1 2 1 】

なお、オンライン処理切替え直後に副系データベース 5 b は、洗い換えをしたデータをバッチ的に複写する間は上記正系サーバー 4 a のデータベース変更メッセージによって更新することができないが、データをバッチ的に複写する時間は比較的短いので、この間に生じたデータベース変更メッセージは一時的に正系の送出用メッセージキュー蓄積手段 1 0 a あるいは副系の受入用メッセージキュー蓄積手段 1 1 b に蓄積され、洗い換えしたデータの複写後に副系データベース 5 b の更新に使用される。

20

【 0 1 2 2 】

図 5 のオンライン処理期間が終了すると、オンライン処理が副系サーバー 4 b に切り替わって図 6 の状態になる。

【 0 1 2 3 】

図 6 に示すように、オンライン処理が副系サーバー 4 b に切り替わると、クライアント 2 a , 2 b , ... , 2 n のいずれかから処理要求が発せられるとゲートウェイ 3 の作動によりそのクライアント (図 6 の例ではクライアント 2 a) は副系サーバー 4 b に接続され、副系のオンラインアプリケーション手段 7 b がオンライン処理を行う。オンラインアプリケーション手段 7 b は、必要に応じて副系データベース 5 b にアクセスし、副系データベース 5 b を変更する場合は、データベースシステム 9 b を介して副系データベース 5 b を変更するとともに、そのデータベース変更メッセージを副系の送出用メッセージキュー蓄積手段 1 0 b と副系メッセージキュー手段 2 3 の双方に送出する。副系メッセージキュー手段 2 3 は、副系サーバー 4 b のオンライン処理期間中のすべてのデータベース変更メッセージを蓄積する。一方、副系の送出用メッセージキュー蓄積手段 1 0 b に送られた副系データベース 5 b に対するデータベース変更メッセージは正系の受入用メッセージキュー蓄積手段 1 1 a に送られ、正系データベース 5 a のバッチ処理終了後に正系データベース 5 a の更新に使用される。

30

【 0 1 2 4 】

一方、正系サーバー 4 a においては、オンライン処理が副系サーバー 4 b に切り替わった直後から、正系データベース 5 a のデータの洗い換えと中間ボリューム記憶装置 2 2 への複写のバッチ処理が行われる (処理 4) 。バッチ処理 4 により、切替え前の処理期間 (正系サーバー 4 a がオンライン処理を行った期間) の終了時点のデータが集約され、データベースファイルの形で中間ボリューム記憶装置 2 2 に複写される。バッチ処理 4 の終了後、正系データベース 5 a と中間ボリューム記憶装置 2 2 の接続は切り離され、中間ボリューム記憶装置 2 2 は、次の切替えによって副系データベース 5 b に複写されるまでアクセスから絶縁された状態になる。

40

【 0 1 2 5 】

処理 4 が終了すると、正系のメッセージキュー起動アプリケーション手段 8 a が起動

50

用アプリケーション制御手段 1 2 a によって起動し、受入用メッセージキュー蓄積手段 1 1 a に蓄積された副系データベース 5 b に対するデータベース変更メッセージを入力し、データベースシステム 9 a を介して正系データベース 5 a を更新する（処理 5）。処理 5 は逐次行われ、正系データベース 5 a は副系データベース 5 b の変更から極めて短い時間経過後更新され、ほぼ常時最新のデータ状態に維持される。これにより、次にオンライン処理が正系サーバー 4 a に切り替えられたときに、最新のデータ状態である正系データベース 5 a を参照することができる。この次に正系サーバー 4 a にオンライン処理が切り替わった状態は図 5 に示すところであり、以降は図 5 と図 6 の状態を交互に繰り返すことになる。上記「24 時間連続稼動オンライン処理システム 2 1」によれば、オンライン処理をするサーバーを切り替えても直ちに最新のデータ状態のデータベースを使用してオンライン処理を行うことができ、かつ、24 時間の切れ目のない運転を行う中で正副両系のデータベースのデータの洗い換え処理、すなわちデータの集約・再登録を行うことができ、不要なデータによるデータ量の膨大化を防止することができるのである。さらに、24 時間連続稼動オンライン処理システム 2 1 は、図 1 ないし図 4 に示した 2 重更新データベースを使用するクライアントサーバーシステム 1 の機能をそのまま有しているため、不測のハードウェア障害やソフトウェア障害が発生したときにも、瞬時に健全なシステムに切り替えられ、システムの回復を実現することができるのである。

【0126】

【発明の効果】

以上の説明から明らかなように、本願発明による 2 重更新を行うデータベースを有するクライアントサーバーシステムによれば、従来、ハードウェアを 2 重化し、障害発生時に障害が発生したシステムのソフトウェアを残るシステムで起動し、起動したシステムを実行することによって障害回復を図るようにしたクライアントサーバーシステムにおいて、データベース等が破損している場合には障害回復することができなかつたのに対し、本願発明のクライアントサーバーシステムは常に 2 重に更新されるデータベースを備えることにより、一方のデータベースならびにその関連ファイルが破損した場合であっても、残る健全なデータベースを参照・更新することにより、直ちにシステムの機能を回復することができる。

【0127】

また、正系サーバーがオンライン処理した所定期間中の正系のデータベースのデータを集約してデータベースファイルの形式で一時的に記憶する中間ボリューム記憶装置と、副系サーバーがオンライン処理した所定期間中の副系のデータベースに対するデータベース変更メッセージを一時的に記憶しておく副系メッセージキュー手段とを備えることにより、正系サーバーと副系サーバーを交互に切替えられ、かつ、正系および副系のデータベースの双方をほぼ常時最新のデータ状態に維持するとともに、定期的に正系サーバーのオンライン処理期間中のデータを集約してデータ量を圧縮できる。これにより、切れ目なく連続運転でき、かつ、ハードウェア障害は言うに及ばず、深刻なソフトウェア障害に対しても直ちに障害回復することができる 2 重更新を行うデータベースを有するクライアントサーバーシステムを得ることができる。

【図面の簡単な説明】

【図 1】本発明の一実施形態による 2 重更新データベースを使用するクライアントサーバーシステムの構成とその通常の処理の流れを示したブロック図。

【図 2】本発明の一実施形態による 2 重更新データベースを使用するクライアントサーバーシステムの構成とハードウェア障害が発生した時の処理の流れを示したブロック図。

【図 3】本発明の一実施形態による 2 重更新データベースを使用するクライアントサーバーシステムの構成とシステムの機能停止を伴わないがデータベース関連ファイルが破損した時の処理の流れを示したブロック図。

【図 4】本発明の一実施形態による 2 重更新データベースを使用するクライアントサーバーシステムの構成とシステムの機能停止を伴うデータベース関連ファイルが破損した時の処理の流れを示したブロック図。

【図5】24時間連続稼動するように構成した本発明の一実施形態による2重更新データベースを使用するクライアントサーバーシステムの構成と、正系サーバーがオンライン処理をしている状態の処理の流れを示したブロック図。

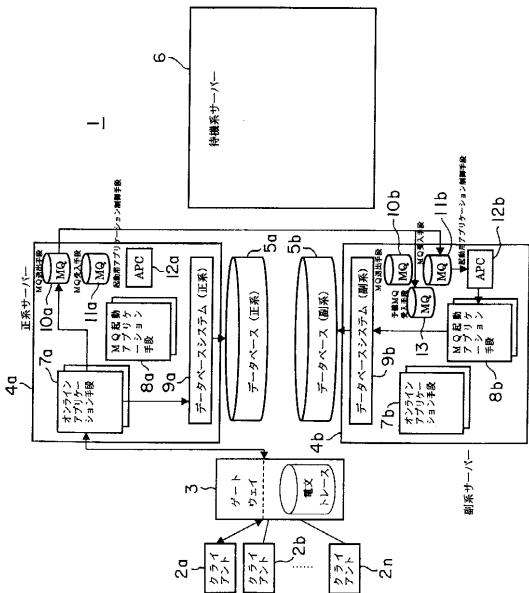
【図6】24時間連続稼動するように構成した本発明の一実施形態による2重更新データベースを使用するクライアントサーバーシステムの構成と、副系サーバーがオンライン処理をしている状態の処理の流れを示したブロック図。

【図7】24時間連続稼動するように構成した本発明の一実施形態による2重更新データベースを使用するクライアントサーバーシステムのオンライン処理の切替えと、データベースの更新処理を示したタイムチャート。

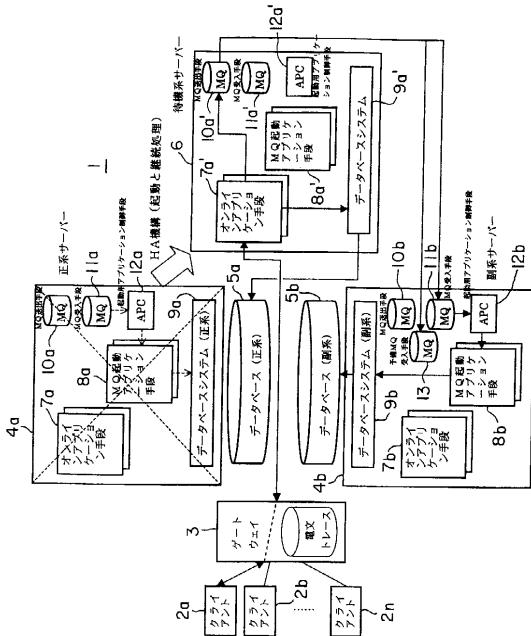
【符号の説明】

- | | | |
|-----|---|----|
| 1 | クライアントサーバーシステム | |
| 2 | クライアント | |
| 3 | ゲートウェイ | |
| 4a | 正系サーバー | |
| 4b | 副系サーバー | |
| 5a | 正系データベース | |
| 5b | 副系データベース | |
| 6 | 待機系サーバー | |
| 7a | 正系のオンラインアプリケーション手段 | |
| 7b | 副系のオンラインアプリケーション手段 | 20 |
| 8a | 正系のメッセージキュー起動アプリケーション手段 | |
| 8b | 副系のメッセージキュー起動アプリケーション手段 | |
| 9a | 正系のデータベースシステム | |
| 9b | 副系のデータベースシステム | |
| 10a | 正系のメッセージキュー送出手段 | |
| 10b | メッセージキュー送出手段 | |
| 11a | 正系のメッセージキュー受入手段 | |
| 11b | 副系のメッセージキュー受入手段 | |
| 12a | 正系の起動用アプリケーション制御手段 | |
| 12b | 副系の起動用アプリケーション制御手段 | 30 |
| 13 | 予備メッセージキュー受入手段 | |
| 21 | 24時間連続稼動する2重更新を行うデータベースを有するクライアントサーバーシステム | |
| 22 | 中間ボリューム記憶装置 | |
| 23 | 副系メッセージキュー手段 | |

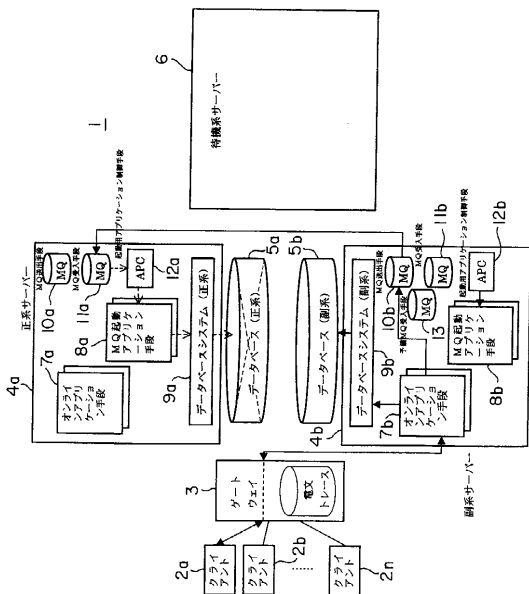
【図 1】



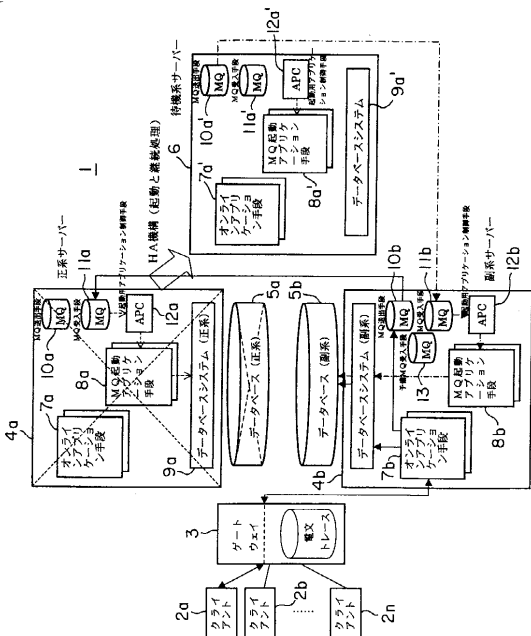
【図 2】



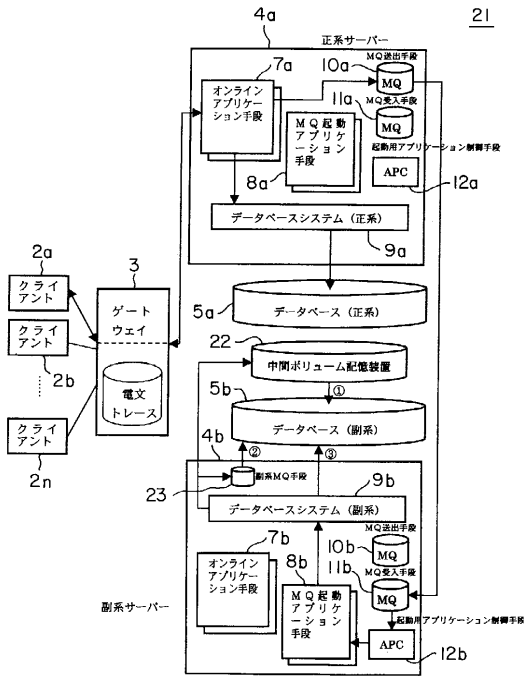
【図 3】



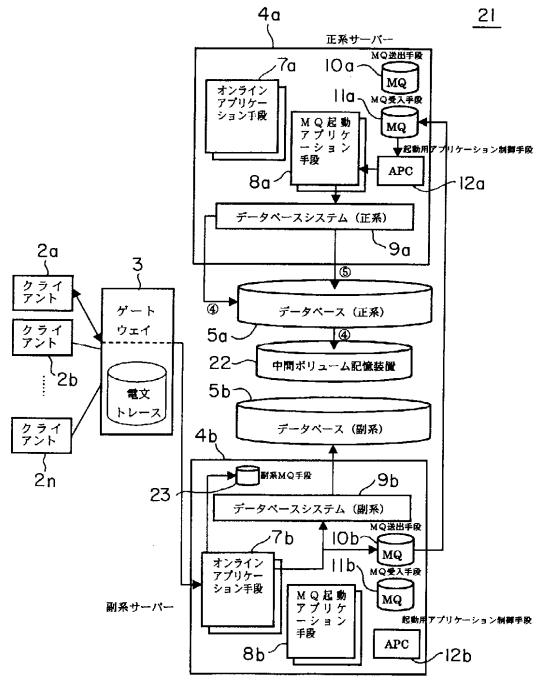
【図 4】



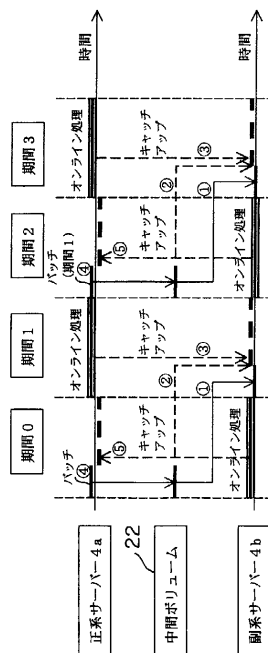
【図5】



【図6】



【図7】



フロントページの続き

(72)発明者 山 藤 健

神奈川県横浜市保土ヶ谷区神戸町134番地 株式会社野村総合研究所内

(72)発明者 高 岸 敏 一

神奈川県横浜市保土ヶ谷区神戸町134番地 エヌ・アール・アイ情報システム株式会社内

合議体

審判長 立川 功

審判官 真木 健彦

審判官 手島 聖治

(56)参考文献 特開平3 - 250257 (JP, A)

特開平10 - 31561 (JP, A)

特開平4 - 299435 (JP, A)

特開平4 - 76742 (JP, A)

特開平3 - 256144 (JP, A)

特開平8 - 106426 (JP, A)

特開平8 - 137814 (JP, A)

特開平8 - 235132 (JP, A)

特開平9 - 62526 (JP, A)

特開平9 - 259096 (JP, A)

特開平10 - 97349 (JP, A)

柳田俊彦, 「障害対策機能の充実進むオフコン」, 日経コンピュータ, 日本, 日経BP社, 1991年7月29日, 第259号, p. 72 - 86

(58)調査した分野(Int.Cl., DB名)

G06F12/00