

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2009-217768

(P2009-217768A)

(43) 公開日 平成21年9月24日(2009.9.24)

(51) Int.Cl.		F I			テーマコード (参考)
<b>G06F 12/00</b>	<b>(2006.01)</b>	G06F 12/00	531A		5B065
<b>G06F 3/06</b>	<b>(2006.01)</b>	G06F 3/06	306K		5B082

審査請求 有 請求項の数 6 O L (全 8 頁)

(21) 出願番号 特願2008-63601 (P2008-63601)  
 (22) 出願日 平成20年3月13日 (2008.3.13)

(71) 出願人 00004237  
 日本電気株式会社  
 東京都港区芝五丁目7番1号  
 (74) 代理人 100096024  
 弁理士 柏原 三枝子  
 (72) 発明者 生形 裕貴  
 東京都港区芝五丁目7番1号 日本電気株式会社内  
 Fターム(参考) 5B065 BA01 CC03 EA19  
 5B082 CA19 DA01 DB02

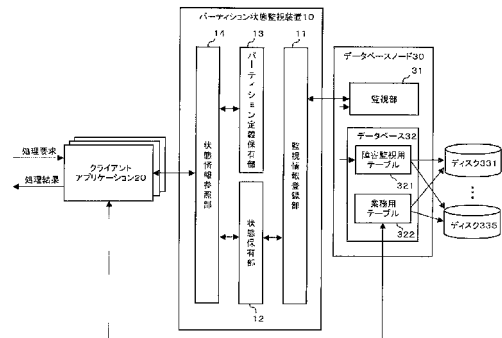
(54) 【発明の名称】 データベースにおけるパーティション障害管理方法およびシステム

(57) 【要約】

【課題】本発明は、パーティショニングにより複数のパーティションに分割されたデータベースにおいて、分割されたパーティション毎の障害情報を管理することにより、パーティション単位のアクセス許可あるいは禁止をすることを課題とする。

【解決手段】複数のパーティションに分割された記憶領域へのアクセス制御方法であって、各パーティションにアクセスして正常か否かを状態保有部に保存する状態検出ステップと、前記記憶領域へのデータ更新要求を受けた場合に、前記状態保有部を参照して要求にかかるパーティションの状態を確認する状態確認ステップと、前記パーティションの状態が正常である場合にのみ前記要求にかかる処理を実行するデータ更新ステップとを含むことを特徴とするアクセス制御方法。

【選択図】 図1



**【特許請求の範囲】****【請求項 1】**

複数のパーティションに分割された記憶領域へのアクセス制御方法であって、各パーティションにアクセスして正常か否かを状態保有部に保存する状態検出ステップと、前記記憶領域へのデータ更新要求を受けた場合に、前記状態保有部を参照して要求にかかるパーティションの状態を確認する状態確認ステップと、前記パーティションの状態が正常である場合にのみ前記要求にかかる処理を実行するデータ更新ステップとを含むことを特徴とするアクセス制御方法。

**【請求項 2】**

請求項 1 に記載の方法において、前記状態検出ステップは定期的に行われ、前記状態保有部には最新の検出状態が保存されることを特徴とするアクセス制御方法。

10

**【請求項 3】**

請求項 1 または 2 に記載の方法において、前記状態確認ステップは、前記記憶領域に格納されるデータがどのパーティションに存在するかを示す対応テーブルを参照して前記要求にかかるデータに対応するパーティションを導出し、当該パーティションの状態を確認することを特徴とするアクセス制御方法。

**【請求項 4】**

複数のパーティションに分割された記憶領域を具える記憶装置において、各パーティションにアクセスして正常か否かを状態保有部に保存する監視手段と、データ更新要求を受けた場合に前記状態保有部を参照して要求にかかるパーティションの状態を確認する状態情報参照手段とを具え、当該パーティションの状態が正常である場合にのみ前記要求にかかる処理を実行することを特徴とする装置。

20

**【請求項 5】**

請求項 4 に記載の装置において、前記監視手段は、定期的に行われ各パーティションへのアクセスを実行して前記状態保有部を更新することを特徴とする装置。

**【請求項 6】**

請求項 4 または 5 に記載の装置において、さらに、前記記憶領域に格納されるデータがどのパーティションに存在するかを示す対応テーブルを具えるとともに、前記状態情報参照手段は、前記対応テーブルを参照して前記更新要求にかかるデータに対応するパーティションを特定し状態確認することを特徴とする装置。

30

**【発明の詳細な説明】****【技術分野】****【0001】**

本発明は、データベースにおけるパーティション障害管理方法およびシステムに関する。特に、パーティショニングにより複数のパーティションに分割されたデータベースにおけるパーティション障害管理方法およびシステムに関する。

**【背景技術】****【0002】**

データベースの可用性、管理性、および問い合わせのレスポンスを向上させるために、1 のデータベーステーブルのデータを複数の論理的なデータの集合であるパーティションに分割するパーティショニングというデータ分割手法が知られている。このパーティショニングにより分割された複数のパーティションは、それぞれ別のディスクあるいは論理ボリュームに割り当てることができ、特定のパーティションを破損した場合でも、他の正常なパーティションへのアクセスやデータの更新等が継続して行うことができる。

40

**【0003】**

しかしながら、テーブルがパーティショニングにより複数のパーティションに分割されている場合でも、このテーブルは、クライアントアプリケーションからは 1 のテーブルとして見えているだけであり、クライアントアプリケーションは、どのようなパーティションが存在するのか認識することができない。このため、クライアントアプリケーションは

50

、パーティションの状態を個別に認識することができず、一部のパーティションに障害が生じても、テーブル全体が正常であるか、あるいはテーブル全体がエラーであるかしか認識できない。パーティショニングの利点は、特定のパーティションに障害が発生しても、他の正常なパーティションが存在する限り、他のパーティションに対しては処理を継続できることである。したがって、この利点を生かすためには、特定のパーティションに障害が発生している場合であっても、他の正常なパーティションへのアクセスを可能にするために、クライアントアプリケーションは、テーブル全体が正常であるとして処理を継続する必要がある。しかしながら、特定のパーティションに障害が発生しているにも拘わらずテーブル全体が正常であるとしてしまうと、障害が発生しているパーティションに複数のトランザクションによる多重アクセスが生じる恐れがある。この場合には、エラーメッセージが繰り返し発生してしまい、他の重要なメッセージが埋もれてしまうという問題がある。

10

**【0004】**

このような問題に対応すべく、障害の発生したデータを含むブロック領域へのアクセスを禁止するファイル閉塞制御方式が開示されている（例えば、特許文献1）。このファイル閉塞制御方式は、障害が発生したデータのアドレスをメモリに記憶しておき、データへアクセス要求があった場合に、アクセスを要求されているデータのアドレスが前記メモリに格納されているか確認し、このアドレスが格納されているときは、当該アドレスへのアクセス要求を拒否する。

20

【特許文献1】特開昭60-258659号公報

【発明の開示】

【発明が解決しようとする課題】

**【0005】**

しかしながら、この方式では、障害のあるデータアドレスへのアクセスを禁止するのみであり、パーティション単位でアクセスを許可あるいは禁止することはできない。また、この方式は、データ領域の障害情報を監視していないため、クライアントアプリケーションが、いずれのメモリへのアクセスが禁止されているのか認識できず、また、データ領域の障害が解消されたことを認識することできない。

**【0006】**

したがって、本発明は、パーティショニングにより複数のパーティションに分割されたデータベースにおいて、分割されたパーティション毎の障害情報を管理することにより、パーティション単位のアクセス許可あるいは禁止することを目的とする。

30

【課題を解決するための手段】

**【0007】**

本発明は、複数のパーティションに分割された記憶領域へのアクセス制御方法であって、各パーティションにアクセスして正常か否かを状態保有部に保存する状態検出ステップと、前記記憶領域へのデータ更新要求を受けた場合に、前記状態保有部を参照して要求にかかるパーティションの状態を確認する状態確認ステップと、前記パーティションの状態が正常である場合にのみ前記要求にかかる処理を実行するデータ更新ステップとを含むことを特徴とする。

40

**【0008】**

また、前記状態検出ステップは定期的に行われ、前記状態保有部には最新の検出状態が保存されることを特徴とする。

**【0009】**

さらに、前記状態確認ステップは、前記記憶領域に格納されるデータがどのパーティションに存在するかを示す対応テーブルを参照して前記要求にかかるデータに対応するパーティションを導出し、当該パーティションの状態を確認することを特徴とする。

【発明の効果】

**【0010】**

本発明は、複数のパーティションに分割された記憶領域へのアクセス制御方法であり、

50

各パーティションにアクセスして正常か否かを状態保有部に保存し、前記記憶領域へのデータ更新要求を受けた場合に、前記記憶領域に格納されるデータがどのパーティションに存在するかを示す対応テーブルを参照して前記要求にかかるデータに対応するパーティションを導出し、当該パーティションの状態を確認し、前記パーティションの状態が正常である場合にのみ前記要求にかかる処理を実行するため、パーティション単位でアクセス許可あるいは禁止することができる。

【0011】

また、本発明は、前記状態検出が定期的に行われ、前記状態保有部には最新の検出状態が保存されるため、ほぼリアルタイムで記憶領域の状態を確認することができ、障害の発生および復旧に迅速に対応できる。

【発明を実施するための最良の形態】

【0012】

以下、本発明を実施するための最良の形態について、添付の図面を参照しながら詳細に説明する。

【0013】

図1は、本発明にかかるデータベースの障害管理システムの全体構成を示す概略図である。本図には、パーティション状態監視装置10と、クライアントアプリケーション20と、データベースノード30とが示されている。このシステムは、例えば、ネットワークに接続されたデータサーバであり、クライアントアプリケーション20が、クライアントPCからデータの処理要求を受け付け、データベースノード30のディスク33に対し処理を行うものである。パーティション状態監視装置10は、データベースノード30から監視情報を収集する監視情報登録部11と、監視情報登録部11が収集した情報を保持する情報保有部12と、予め定義されたパーティション情報を保持するパーティション定義保有部13と、クライアントアプリケーション20がパーティションの状態を参照するための状態情報参照部14とを備えている。

【0014】

図2は、データベースノード30の監視部31をより詳細に示した図である。データベースノード30は、監視部31と、データベース32とを備えており、このデータベース32は、障害監視用テーブル321と、業務用テーブル322とを備えている。監視部31は、データベース32の障害監視用テーブル321を参照してパーティション障害が発生しているか判断する判断手段312と、当該判断手段312の判断結果を、パーティション状態監視装置10の監視情報登録部11に送信する通信手段311とを備えている。

【0015】

図3は、データベースノード30のデータベース32が備える障害監視用テーブル321および業務用テーブル322と、ディスク33との関係を示す図である。本図に示すように、本実施例では5つのディスクを使用しているため、これらのディスクに対応するように障害監視用テーブル321を5つのパーティション3211乃至3215に分割する。本実施例では、パーティションを物理的なディスク単位で割り当てているが、代替的に論理ボリューム単位で割り当ててもよい。また、業務用テーブル(A)322aは、4つのディスク331乃至334に割り当てられ、業務用テーブル(B)322bは、2つのディスク334および335に割り当てられている。

【0016】

図4は、障害監視用テーブル321の定義例を示しており、図5は、この定義例による障害監視用テーブル321の初期状態を示している。この定義例では、ID列を主キーに設定し、IDが重複しないようにしている。また、リストパーティションにより、ID1乃至5の各レコードが、障害監視用テーブル321のパーティション3211乃至3215と一意に対応するように定義されている。これらのパーティションはそれぞれ、ディスク331乃至335に対応している。これにより、ID1乃至5の各レコードにSQL UPDATA文等を発行し、その結果を確認することにより、パーティション単位で障害の有無を確認することができる。

10

20

30

40

50

## 【0017】

次に、図6を参照して、本発明の処理動作について説明する。初めに、データベースノード30内の監視部31の判断手段312が、障害監視用テーブル321にSQL SELECT文を発行し(A1)、障害監視用テーブル321に格納されたIDのリスト(ID1乃至5)を取得する(A2)。これらのIDが障害監視用テーブル321の各パーティションおよびディスクに対応する。次に、監視部31の判断手段312は、パーティションの障害の有無を確認すべく、各IDに対応するレコードにSQL UPDATE文を発行し(A3)、更新結果を受け取る(A4)。ここで、SQL文が正常終了した場合、該当するパーティションは正常であると判断し、SQL文が異常終了した場合、該当するパーティションは異常であると判断する。次に、監視部31の通信手段311が、この更新結果をパーティション状態監視装置10の監視情報登録部11に送信し(A5)、当該監視情報登録部11は、この更新結果を状態保有部12に登録する(A6)。図7は、この状態保有部12の一例を示す図である。状態保有部12には、これらの更新結果がIDおよびディスク毎に保持される。データベースノード30の監視部31は、これらの処理を定期的に、例えば1秒おきに繰り返す。これにより、パーティションの状態が定期的に更新され、パーティション毎の状態をほぼリアルタイムで確認することができる。

10

## 【0018】

次に、図8乃至10を参照して、パーティション定義保有部13について説明する。パーティション定義保有部13は、これらの図に示すように、テーブル/パーティションキー対応テーブル131と、業務テーブルA用のキー値/ディスク名対応テーブル132aと、業務テーブルB用のキー値/ディスク名対応テーブル132bとを備えている。テーブル/パーティションキー対応テーブル131は、図8に示すように、業務テーブルAおよびBのパーティションキーが登録されている。業務テーブルA用のキー値/ディスク名対応テーブル132aは、図9に示すように、パーティションキー「date」の範囲と、これに関連するディスク名とが登録されている。業務テーブルB用のキー値/ディスク名対応テーブル132bは、図10に示すように、パーティションキー「ID」の範囲と、これに関連するディスク名とが登録されている。

20

## 【0019】

次に図11を参照して、クライアントアプリケーション20が、業務テーブルを更新する処理について説明する。以下の説明では、更新の対象の業務テーブルを「業務テーブルA」とし、キー値を「2007.1.1」とする。

30

## 【0020】

初めに、クライアントアプリケーション20が、クライアントPC(図示せず)から業務用テーブルの更新要求を受信すると(B1)、パーティション状態監視装置10の状態情報参照部14に、更新の対象であるテーブル名「業務用テーブルA」および更新する情報を送信する(B2)。次に、状態情報参照部14は、パーティション定義保有部13のテーブル/パーティションキー対応テーブル131を参照して、受信したテーブル名「業務用テーブルA」に関連するパーティションキー「date」を取得する(B3)。次に、状態情報参照部14は、取得したパーティションキー「date」と当該キーの値「2007.1.1」を基に、パーティション保有部13の業務テーブルA用のキー値/ディスク名対応テーブル132aを参照して、該当するディスク名「ディスク333」を取得する(B4)。状態情報参照部14は、図7に示す状態保有部12を参照して、「ディスク333」の状態情報を取得し(B5)、この状態情報をクライアントアプリケーション20に送信する(B6)。クライアントアプリケーション20は、状態情報が正常である場合にのみデータベースの更新処理を実施する(B7)。このように、クライアントアプリケーション20は、データベース更新処理を行う前にパーティションの状態を確認するため、不必要なデータベースの更新を回避することができる。

40

## 【0021】

ここでは、更新対象のテーブルが1つの場合の処理についてのみ説明しているが、更新対象のテーブルが複数ある場合にはこれらの処理を繰り返し行う。この場合も同様に、更

50

新対象であるパーティションが正常である場合にのみ更新処理を行う。

【0022】

本発明の好適な実施例を示し説明したが、本発明の目的または特徴から逸脱しない範囲で様々な変更を加えることができる。

【産業上の利用可能性】

【0023】

本発明は、データベースのパーティショニングを利用しているシステム全般に用いることができる。特に、トランザクション処理件数の多い大規模データベースを有するオンラインシステムに有用である。

【図面の簡単な説明】

【0024】

【図1】図1は、本発明にかかるデータベースの障害管理システムの全体構成を示す概略図である。

【図2】図2は、データベースノード30の監視部31の詳細図である。

【図3】図3は、データベース32の障害監視用テーブル321および業務用テーブル322と、ディスク33とを示す図である。

【図4】図4は、障害監視用テーブル321の定義例を示す図である。

【図5】図5は、図4の定義例による障害監視用テーブル321の初期状態を示している。

【図6】図6は、本発明の処理動作を示す図である。

【図7】図7は、パーティション状態監視装置10の状態保有部12を示す図である。

【図8】図8は、業務テーブルAおよびBのパーティションキーを管理するテーブル/パーティションキー対応テーブル131を示す図である。

【図9】図9は、パーティションキー値「date」の範囲と、これに関連するディスク名を管理する業務テーブルA用のキー値/ディスク名対応テーブルを示す図である。

【図10】図10は、パーティションキー値「ID」の範囲と、これに関連するディスク名を管理する業務テーブルB用のキー値/ディスク名対応テーブルを示す図である。

【図11】図11は、クライアントアプリケーション20が、業務テーブルを更新する処理を示す図である。

【符号の説明】

【0025】

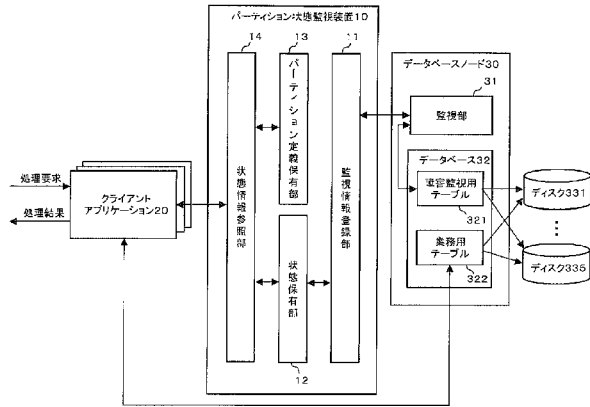
- 10 パーティション状態監視装置
- 20 クライアントアプリケーション
- 30 データベースノード

10

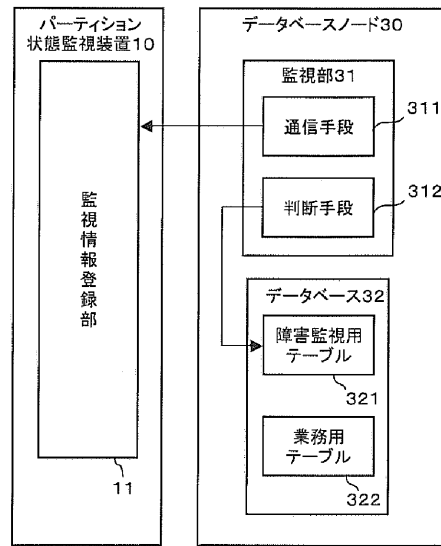
20

30

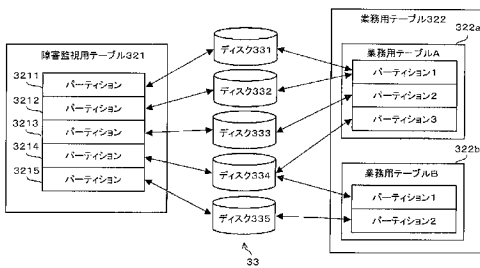
【 図 1 】



【 図 2 】



【 図 3 】



【 図 5 】

ID	DUMMY
1	1
2	1
3	1
4	1
5	1

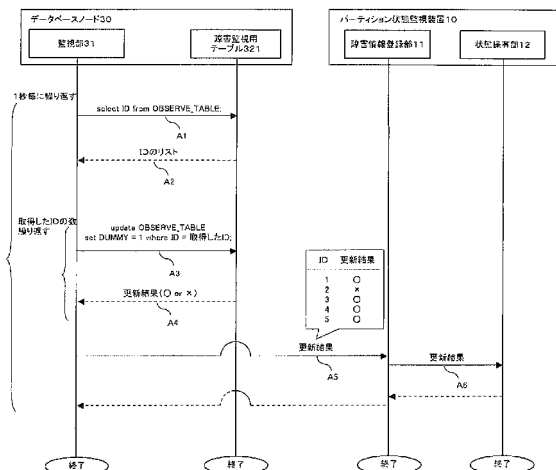
【 図 4 】

```

障害監視用テーブル321の定義
create table OBSERVER_TABLE
(
  ID number ( 2 ),
  dummy number ( 2 ),
  primary key ( ID )
)
partition by list ( ID )
(
  partition part1 values ( 1 ) tablespace ディスク331に対応するテーブルスペース,
  partition part2 values ( 2 ) tablespace ディスク332に対応するテーブルスペース,
  partition part3 values ( 3 ) tablespace ディスク333に対応するテーブルスペース,
  partition part4 values ( 4 ) tablespace ディスク334に対応するテーブルスペース,
  partition part5 values ( 5 ) tablespace ディスク335に対応するテーブルスペース
);

```

【 図 6 】



【 図 7 】

ID	ディスク名	状態
1	ディスク331	OK
2	ディスク332	NG
3	ディスク333	OK
4	ディスク334	OK
5	ディスク335	OK

【 図 1 0 】

パーティションNo.	キー範囲	ディスク名
1	ID < 10000	ディスク334
2	10000 <= ID	ディスク335

【 図 8 】

テーブル名	パーティションキー
業務テーブルA	date
業務テーブルB	ID

【 図 9 】

パーティションNo.	キー範囲	ディスク名
1	2006.4.1 <= date date < 2006.10.1	ディスク331 ディスク332
2	2006.10.1 <= date date < 2007.4.1	ディスク333
3	2007.4.1 <= date date < 2007.10.1	ディスク334

【 図 1 1 】

