保証レベルを導入した主メモリ常駐DBのリカバリ方式

3Q－7

日高東潮　小林伸幸　板倉一郎
ＮＴＴ情報通信研究所

1　はじめに

高度通信サービスで使用されるデータベース（以下DB）では、高速性が重要である。その高速性を満たすDBの一つに、検索時にI/Oの発生しない主メモリ常駐DB[1](以下MMDB)がある。

一般にMMDBではDBのバックアップ(以下BK)を定期的に二次記憶に取得し、障害に備えている[2][3]。高度通信サービス用DBではさらに高信頼性を確保するため、BKを重複して持つ冗長構成を取ることが多く、二次記憶のコストを増大させている。

従来の情報処理用DBでは、全データのBK取得を行なうことで、障害回復時に障害直前の値を保証する。しかし、高度通信サービス分野ではデータ特性を考えた場合、必ずしも障害直前の値への復帰が必要ないデータが存在する。この特性を利用し、障害直前の値を保証すべきデータのみBKの取得を行なうことで、必要な二次記憶の容量を抑え、コストの削減を図ることが可能となる。

本稿では高度通信サービスのデータ特性を考慮した、部分的なBK取得による障害回復方式について検討し、その実現方式を提案する。

2　保証レベルに基づく提案方式

高度通信サービス用DBでは動作無中断が重要となるため、BKの取得をトランザクション処理を中断することなく行う必要がある。その一手法として、Fuzzy Check Point方式(以下FCP方式)がある[1]。これはトランザクション処理とは非同期にCheck Point取得(以下CP取得)を二次記憶上に行なう方式である。本方式では、データの整合性を保持するため、CPと共にCP取得期間中の更新に対する更新前後情報が二次記憶上に必要となる。

障害回復時には、CPのメモリ上へのロードと更新前後情報の反映によりデータ間の整合性がとれている状態のDBを復元し、さらにCP取得期間外の更新後情報によるロールフォワードにて障害直前の値に復帰させる。

現状の障害回復方式では、全DBのBKを取得する。それに対し、本稿ではDBの一部データ(障害直前の値への復帰を保証すべきデータ、以下保証データ)のみのBKを取得する方法を検討することにより、バックアップの二次記憶上に必要とする領域を削減する。

本稿では、保証データの範囲を”保証レベル”と呼ぶことにする。また、高度通信サービスにおいては、サービスの仕様上レコードとカラムを保証・非保証の単位とする。このため、本稿では保証レベルとして、レコードとカラムをする。障害回復時、非保証データは通信処理を考慮した固定値による初期化を行なうことで障害回復を行なうことにする。本稿で提案する方式のイメージを図1に示す。

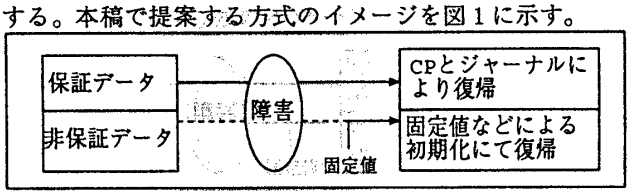

図1：本稿で提案する障害回復方式

次節でそれぞれの保証レベルについて、実現方式を検討する。

3　実現方式の検討

前述の2つの保証レベルに対する障害回復方式の実現上、FCP方式とジャーナル取得方式について実現方式を検討する必要がある。以下に本稿における方式を示す。

・FCP方式

各保証レベルにおける、障害回復を保証するデータ範囲に対してのみ、CPの取得を行なう。

・ジャーナル取得方式

更新対象のデータが保証データであるか否かを判定し、保証データである時だけジャーナルを取得する。

A Recovery Method of Main Memory Resident Database based on Guarantee Level.
Toshio Hitaka, Nobuyuki Kobayashi, Ichiro Itakura
NTT Information and Communication Systems Laboratories
1-2356 Take,Yokosuka,Kanagawa 238-03 Japan

上記方式を実現するにあたり、保証データのメモリ上の配置が、CP 取得やメモリ上への CP のロードなどの処理時間に影響を及ぼす。

レコード単位の保証レベルの場合、メモリ上の保証レコードの配置には以下の２通りの方式が考えられる。

- 方式 1-1: 保証・非保証データを混在させる方式
- 方式 1-2: 保証・非保証データを分離させる方式

各方式のイメージを図２に示す。

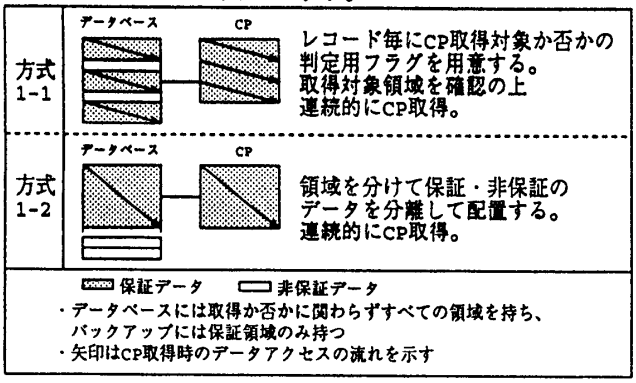

図２：レコードに基づくFCP方式

カラム単位の保証レベルの場合、メモリ上の保証カラムの配置には以下の３通りの方式が考えられる。

- 方式 2-1: 保証・非保証データを混在させる方式
- 方式 2-2: レコード内で保証・非保証データを分離させる方式
- 方式 2-3: テーブル内で保証・非保証データを分離させる方式

各方式のイメージを、図３に示す。

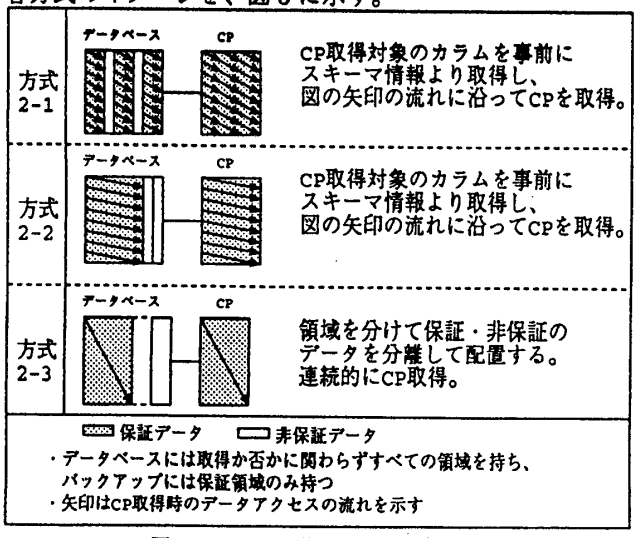

図３：カラムに基づくFCP方式

4 考察

前述の各方式を、以下の点から定性評価した結果を表１に示す。

- 二次記憶媒体使用量
- 処理時間

(本来処理 (select,insert 等)・FCP 取得処理・障害回復処理)

表1: FCP方式に対する、本方式の適用範囲

方式	2次記憶使用量	処理時間	判定理由
1-1	○	×	CP取得時のデータアクセス回数が多いため
1-2	○	○	連続的なCP取得により、データアクセス回数小のため
2-1	○	×	CP取得時のデータアクセス回数が多いため
2-2	○	×	CP取得時のデータアクセス回数が多いため
2-3	○	○	連続的なCP取得により、データアクセス回数小のため

CP 取得領域のデータ配置の連続性は、CP 取得時の処理時間に大きく影響し、連続であることが不可欠である。その観点から、結論として、方式 1-2、方式 2-3 が有効であることが分かる。

5 まとめ

今回、高度通信サービス用 DB において、データの障害回復を部分的に保証する方式として保証レベルを規定し、それに基づく障害回復方式を提案した。また、該方式の定性評価により、バックアップとジャーナル取得に必要な二次記憶の容量を削減することが可能であることを示した。

今後プロトタイプによる定量評価を行なう予定である。

参考文献

[1]Kenneth Salem, Hector Garcia-Molina: 'SYSTEM M: A Transaction Processing Testbed for Memory Resident Data',IEEE TRANSACTIONS ON KNOWLEDGE AND DATAENGINEERING VOL.2 NO.1 MARCH 1990

[2]J.Gray 他:FAULT TOLERANT SYSTEM, Mc.Grawhill,198

[3]J.Gray 他:ON-LINE TERANSACTION PROCESSING SYSTEM Mc.Grawhill,1991