

2フェーズオンラインDB再構成方式*

7W-7

堀口 恭太郎 小林伸幸
NTT情報通信研究所

1. はじめに

24時間無中断で稼働するネットワークの構成要素としてのデータベース管理システム(DBMS)では、サービスを継続しながらデータベース(DB)再構成を行うオンラインDB再構成が必要である。ここでの"オンライン"はトランザクション処理が行われている(処理を停止しない)ことを意味する。

現状、オンラインDB保守としてはおもにデータのロード/アンロード、再編成が実現されている。しかしDBのスキーマを変更する再構成や、データ実体のファイル内での移動を伴う再編成の処理は、一般的にサービスをオフライン状態にして行う必要がある。

ここでは、オンライン状態でDB再構成処理を行うための方法として2フェーズDB再構成方式を提案する。また、メインメモリデータベースで性能を向上する方法としてロックヒストリを用いた更新同期方式を提案する。

2. オンラインDB再構成方式

2.1 基本方式

オンライン再構成は基本的にはサービス/スタンバイの2台のシステムを用いて図1のような方式で実現できる[1]。

しかし、冗長系の中の1台のシステムに再構成を行わせるこの方式では冗長系全体の耐障害性の低下を招くことになる。

2.2 2フェーズオンラインDB再構成

そのような耐障害性の低下を防ぐためには、単独のシステムでオンラインDB再構成を行う方法が必要となるが、これは基本的には図1と同様の方法を1システム内で行えばよい。(図2)

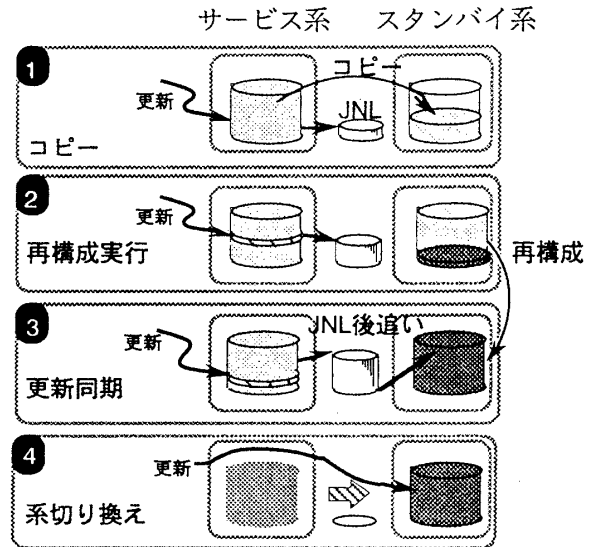


図1. 二重化システムでのオンラインDB再構成

二重化システムでの再構成は、サービス系の負荷を下げるためコピーと再構成を分けて行っていた。しかし、1システムでの場合はトータルの処理量を減らす方が性能向上が望めるため、コピーと再構成を同時に行い処理をコピー・再構成(フェーズ1)→更新同期(フェーズ2)の2フェーズとする。

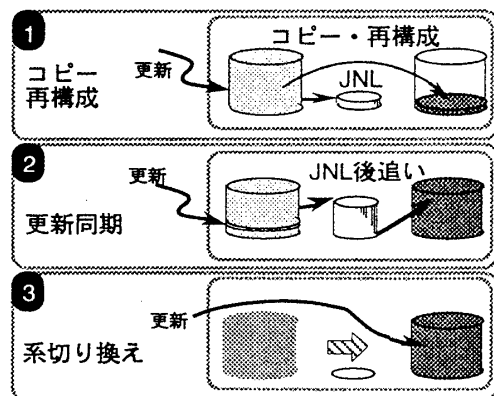


図2. 単独システムでのオンラインDB再構成

* 2 phase on-line database restructuring.

Kyotaro HORIGUCHI, Nobuyuki KOBAYASHI
NTT Information and Communication Systems Laboratories
1-2356 Take Yokosuka-Shi Kanagawa 238-03, Japan

2.3 問題点

2次記憶ベースのデータベースシステムでは、この方式でオンライン再構成は実現可能であるが、メインメモリデータベースシステムでは、トランザクション処理時間が短く(高スループット)、記憶領域のビット単価が比較的高い。そのため更新同期にジャーナルそのものを用いるこの方法では以下のような点が問題となってくる

- ジャーナルを一旦2次記憶に格納し、そこから必要なものを読み出すという方法では、I/Oが増加し、再構成処理速度の低下につながる。
- ジャーナルをメモリに格納すると、トラフィックが高い時には、ジャーナルが大量に発生し、その格納のために大量の記憶領域が必要となる。

2.4 ロックヒストリを用いた更新同期方式

2フェーズDB再構成では再構成前後のDBが同じシステム内にあるため、更新同期のためには更新されたレコードが特定できれば、旧テーブルから更新後のレコードを取り出してくることができる。これによりジャーナル情報は基本的に不要となる。

そこで処理が高速で、かつ必要な記憶領域が少なくすむようにするために、ここでは再構成処理中の排他制御情報として、更新されたレコードのレコードID(RID)のみをメモリ上に保存して(ロックヒストリ)それを用いて更新同期を行う。(図3,図4)

方式の概略は以下のようなになる。

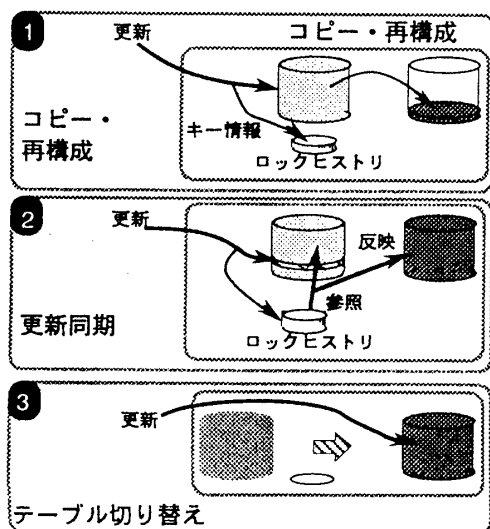


図3. ロックヒストリによる更新同期を用いた2フェーズ再構成

(1) フェーズ1:コピー・再構成

- 現在サービス中のDBとは別に再構成作業用のテーブル(新テーブル)を作成する。
- サービス中のDBの再構成対象テーブル(旧テーブル)のコピーをしながら再構成を行う。
- この間に旧テーブルが更新されたときはそのRIDをロックヒストリとして蓄積する(図4-a)。

(2) フェーズ2:更新同期

- フェーズ1が終わったら、ロックヒストリを参照し(図4-b)、更新されたレコードを旧テーブルから検索し(図4-c)、これを新テーブルに対して反映する(図4-d)(更新同期)。
- これをロックヒストリが空になるまで行う。

(3) テーブル切り替え

更新の同期が取れたら、アクセスを新テーブルに切り替える。

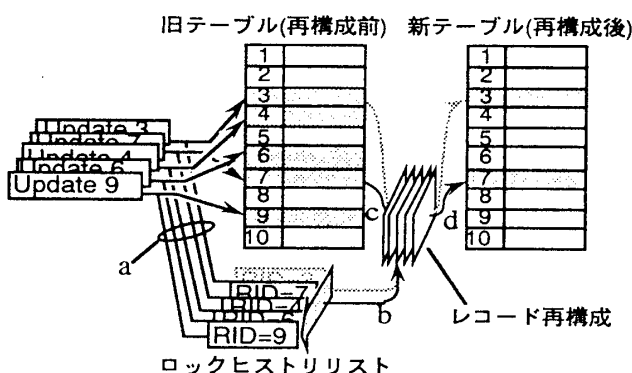


図4. ロックヒストリを用いた更新同期方法

3. まとめ

1システム上でオンライン再構成を実現し、さらに更新同期処理のためにジャーナルのかわりに対象テーブルのロックヒストリを用いる2フェーズDB再構成方式を提案した。この方式はメインメモリデータベースシステムに適し、高速でかつ省スペースなオンラインDB再構成方式を実現する。

参考文献

- 1) 梅本佳宏他:データベース・オンライン保守の方式, 情報処理学会第46回全国大会講演論文集 4-181 (1992).