

# 分散OLTPシステムにおける ディレイドオンライン制御

2T-5

香川 弘一\* , 滝本 秀明\* , 関田 武志\*\*

\* (株) 東芝 府中工場 \*\*東芝アドバンスドシステム (株)

## 1. はじめに

オンライントランザクション (OLTP) システムでは、トランザクション (以降TX) は通常発生時点で処理される。しかし、システム形態の、集中型から分散型への変化に伴い、分散処理を行う一部の計算機や、通信回線の障害が、システム全体の機能停止へとつながる危険性が高くなった。そこで、送信先の計算機が障害等の理由により、TX要求を送信出来ない場合、時間的に遅延させて実行を行うディレイドオンライン処理が必要となる。今回、UNIXサーバー上の当社独自のTPモニターであるTX/ATPS (I-7611) に、下記の特徴を持つディレイドオンライン機構を実現したので報告する。

- ファイルキューによるメモリ資源の有効利用。
- ファイル型D/Cバッファプールによる高速処理。
- メッセージ通番による順序性の保証。

## 2. 設計方針

TX/ATPSでは、リカバリ可能なメモリキュー<sup>[1]</sup>を介してリアルタイムにTX処理が行われる。しかしディレイドオンライン処理では、すぐにメモリの解放が行われずメモリ資源を圧迫する。そこで、新たなメッセージキューとしてファイルキューを導入した。ただし、その機構はユーザから隠べし、宛先の状態により、自動的に切り替え処理を行う事とした。また、TXメッセージの順序性を保証し、ファイルキューの構造や、プロセスの機能分担を工夫し、大規模なオンラインシステムに耐えうる機構を実現する事とした。

## 3. システム構成と処理の流れ

TX/ATPSは図1に示すようなサブシステムで構成される。ディレイドオンライン機構は、メッセージ制御部のファイルキュー制御機構により実現される。

システム内では、TXクライアント (TC) とTXサーバ (TS) 間で、TXメッセージの送受信により処理が進められる。この時、メッセージ制御部は各宛先毎が持つメッセージキューにより、スケジューリング管理を行う。宛先キューは、オンライン/オフライン/閉塞の3つの状態を持ち、宛先がオンライン状態でない場合、ファイルキューへの格納処理が自動的に行われる。またオンライン状態に遷移するタイミングでファイルキューから取り出され、スケジューリングされる。

宛先の状態の変更は、下記の手段が提供される。

- 状態変更コマンドによる遷移
- 状態変更ライブラリによる遷移
- 障害などのイベントに連動した遷移。

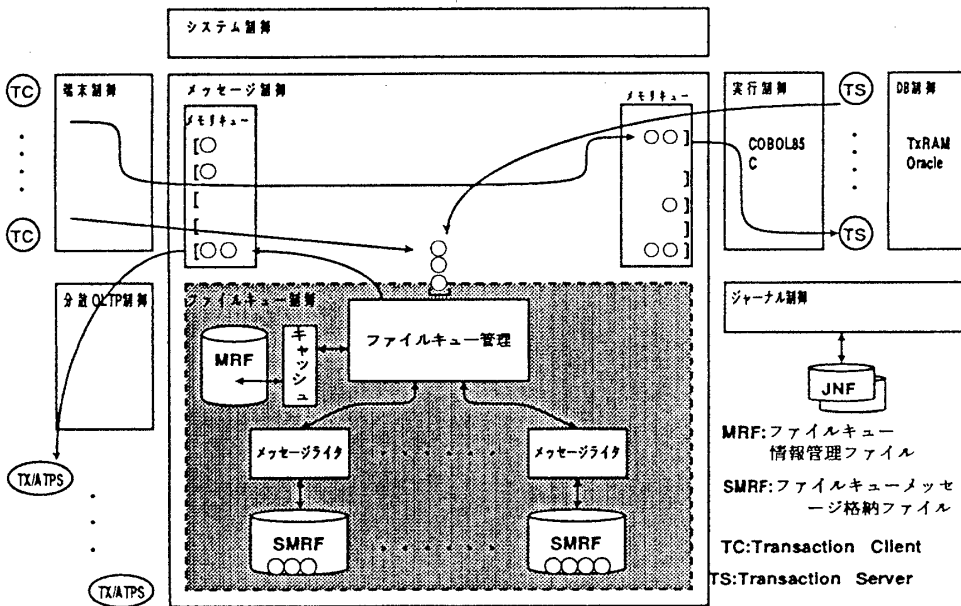


図1 ATPSのシステム構成

Delayed online control for Distributed OLTP  
Koichi Kagawa, Hideaki Takimoto, TOSHIBA CORP.  
Takeshi Sekita, TOSHIBA ADVANCED SYSTEM CORP.

4. ディレイドオンライン機能のための要素技術

(1) ファイルキュー制御

環境設定時に、宛先毎に必要な情報管理ブロックをファイル上に確保する。(図2) 格納要求時、TXメッセージを物理ファイルに書き込み後、メッセージジャーナルに対するクリーンポイントジャーナルを取得する。ファイルキューの管理情報は、毎回でなく定期的に物理ファイルへの書き込み処理を行う。

(2) 抑止制御

資源の使い切りを防止するため、抑止機構を持つ。

●最大滞留メッセージ制御

メモリキューにある一定数を越えたメッセージが滞留した場合、それ以降のメッセージをファイルキューに格納し、メモリ資源の使い切りを防止。

●最大格納数制御

ファイルキューの長さに制限を設け、ファイル資源の使い切りを防止する。ファイルキューを溢れたTXメッセージは、トランザクションのキャンセルが発生した物として扱われる。

(3) ファイル型DCバッファプール制御

環境設定時、あらかじめ必要な量のディスク領域を、異なるサイズの複数のブロックに分割してプールとして確保する。(図2) ファイルキューへの格納要求時、プール中から最適なサイズの未使用領域が割り当てられる。またファイルキューから取り出された場合は、元のプールに戻される。

●複数ディスク分散配置機能

●しきい値による自動拡張機能

(4) 並列処理

1つのファイルキュー管理タスク(MRM)の下に、ディスクI/Oを司るメッセージライター(MWTR)を複数配置し、複数の要求を並列処理できる機構とする。

(5) ダウンリカバリ処理

システム障害発生時は、ジャーナルファイル(JNF)中のメッセージジャーナルと、物理ディスク上に残ったファイルキュー(MRF, SMRF)上のメッセージにより、障害復旧処理

が行われる。ファイルキューの制御部の情報もファイルに記憶されているため、システム障害発生後も、高速にファイルキューの復元処理を行うことができる。

ただし制御領域の書き出しを、即座に行わないため、システム障害発生時、制御情報が失われる可能性がある。この場合は、タイムスタンプによりファイル中の情報が正しいか判定し、不整合の場合は、格納されたTXメッセージから制御情報の復元を行う。

(6) メッセージ通番制御

各宛先毎にメッセージ通番器を持ち、通番をTXメッセージに付加し、メッセージの送受信やリカバリ、ファイルキュー上での順序性を保証する。特に、分散トランザクション制御では、ネットワークをまたがったTXメッセージの重送や脱送を防止する。

5. おわりに

本報告では、分散OLTPシステムでは、全てのトランザクションを発生時点で処理する事ができない場合があるという新たな問題点に着目し、ファイル型DCバッファプール制御、メッセージライター方式と言った技術により、大規模なオンラインシステムでの利用に耐えるディレイドオンライン機構を実現した。また、メッセージ通番機構により、分散OLTPでのネットワークをまたがったTXメッセージの順序性、信頼性を保証した。

参考文献

[1]香川 他:オンラインシステムのメッセージリカバリ手法

情報処理第42回全国大会(1990)

