オンラインシステムのメッセージリカバリ手法

4L−11

香川　弘一*　，　滝本　秀明*　，　飯島　秀之**

*（株）東芝　府中工場　　**日本システム（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、トランザクションのデータは１つのメッセージ単位として管理されている。そしてＯＬＴＰの４特性（原子性、一貫性、独立性、耐久性）から、たとえシステムになんらかの障害が発生したとしても、それらのメッセージの紛失などが絶対にあってはならない。

ミニコンピュータ（ＤＳ６５００）上の「トランザクション管理」において、以下の特徴を持つトランザクションメッセージの復元（リカバリ）機能を実現したので報告する。

●メッセージとデータベースの一貫性の保証
●機能分担による処理の高速化
●メッセージ通番による順序性の保証

2．障害の分析と設計方針

まずはじめに、計算機システムで発生する障害を分析し、それらの障害に対しどの様なリカバリを行なうかを決定した。

障害は大きく分類すると以下の２つに分類できる。

(a)端末障害：局所的障害
　　通信障害、電源断等による端末の障害
(b)システム障害：大域的障害
　　電源断等による計算機システムの停止

以上の障害に対し、下記の設計方針を立てた。

●明確な機能分担を行い、処理の並列化によりリカバリを高速化する。
●リカバリ準備（ジャーナル取得など）によるオーバヘッドを最小限にし、通常のトランザクション処理の高速性を実現する。

3．機能分類

トランザクション管理のリカバリ管理は、下記の機能で構成され、相互の機能の連携によりリカバリの高速化および信頼性を実現している。

●システムリカバリ管理
●メッセージリカバリ管理
●メッセージリカバリファイル管理
●メッセージ管理
●システム状態管理

これらの機能単位の関係を図−１に示す。

4．メッセージリカバリのための要素技術

（１）メッセージリカバリファイル（ＭＲＦ）を導入し、端末障害時のリカバリを高速化する。

●この機能は、端末障害時のみ動作することで性能低下を防止する。
●ＭＲＦの各端末毎のメッセージの格納場所は固定であり、高速に格納／取り出しが行なえる。
●メッセージの圧縮を行いＭＲＦに格納する

（２）ジャーナルファイルからのリカバリ処理を複数のタスクで並列に行ない、システム障害時のリカバリ（ダウンリカバリ）を高速化する。

（３）メッセージ通番を入力／出力メッセージに付加し、メッセージの送受信やリカバリ時の順序性を保証する。

（４）システム稼働中に、各トランザクションの開始前に入力メッセージジャーナル（ＩＭＪ）を取得し、トランザクションのコミット時に出力メッセージジャーナル（ＯＭＪ）を取得する。ジャーナルデータは必要な情報のみ残して圧縮して格納する。

（５）ジャーナル取得による性能低下の防止のため以下の２点を実現した。

●複数トランザクションのジャーナルをまとめて書き出してＩ／Ｏ回数を減らす。（ジャーナルの一括書き出し）
●ＯＭＪ取得時に、ＩＭＪのクリーンポイントジャーナル（ＣＬＰＪ）とデータベースの更新後イメージジャーナル（ＡＩＪ）の同時書き込みを行う。

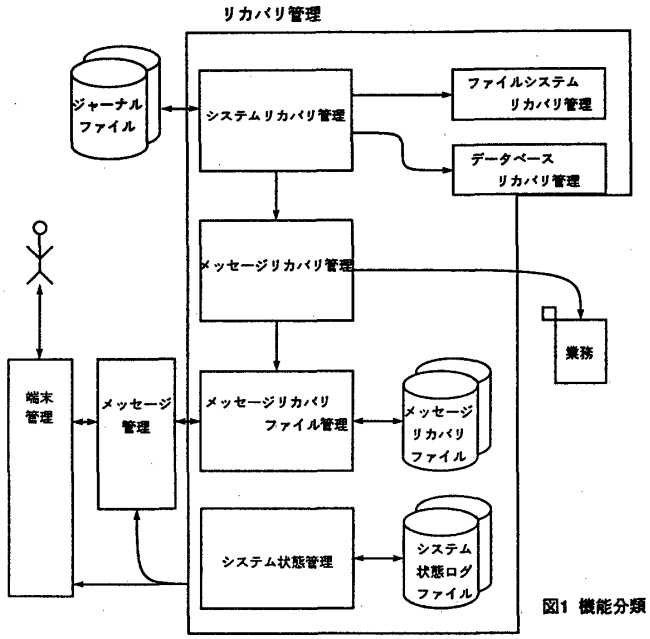

図1 機能分類

Message Recovering Methods for OLTP on middle range Computers.

Kouichi Kagawa.　Hideaki Takimoto.
TOSHIBA CORPORATION
Hideyuki Iijima　NIHON SYSTEM CORPORATION

5．各障害時のリカバリ処理

5．1　端末障害時のメッセージリカバリ

端末障害のような局所的な障害が発生した場合、システム全体としては稼働を続け、その端末から投入されたトランザクション入力メッセージは業務プログラムに渡されて処理を行なう。送信できない入／出力メッセージは、メッセージリカバリファイル管理に渡され、ファイルに一時的に保存される。端末の復旧後は、メッセージリカバリ管理に要求が出され、その端末に対するメッセージがリカバリされる。（図２）

このような機構を設けることにより、局所的な障害に対するリカバリを高速化できる。

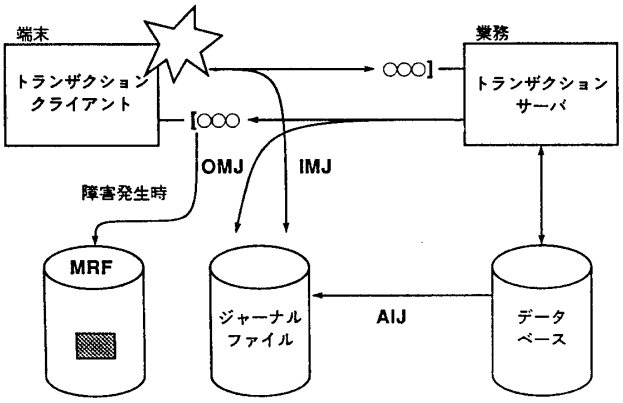

図2 端末障害時の流れ

5．2　システム障害時のメッセージリカバリ

システム障害が発生した場合は、ジャーナルファイルからメッセージのリカバリを行なわなくてはならない。しかし、通常ジャーナルファイルは膨大であり、メッセージ以外にもファイル、データベース等の情報を含んでいる。そこで以下の機能分担により、リカバリの高速化を実現する。

1）システムリカバリ管理

システムリカバリ管理はシステムの再起動時に、ジャーナルファイルを直接解析し、リカバリすべきメッセージやデータを捜し出す。そして各データをメッセージリカバリ管理、ファイルシステム管理、データベースリカバリ管理へ分配する。

2）メッセージリカバリ管理

システムリカバリ管理からパイプを通して受信したデータを、メッセージとして復元する。そして入力メッセージは業務プログラムへトランザクションの再投入を行なう。出力メッセージは、メッセージリカバリファイル管理に渡して、メッセージリカバリファイルに一時的に格納する。

3）メッセージリカバリファイル管理

送信できないリカバリメッセージをリカバリファイルへ格納する。また端末の復旧、再結合が完了した場合には取り出し要求に従い、リカバリファイルから取り出して再送信する。（図３）

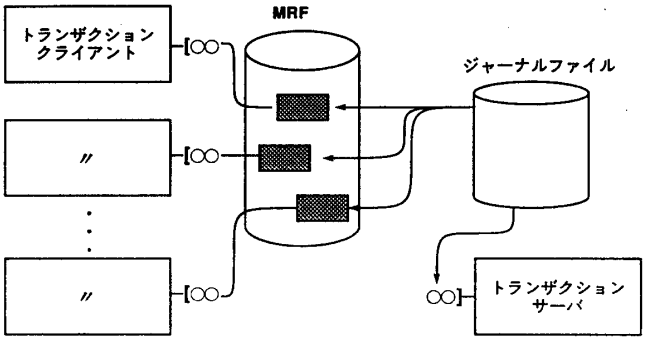

図3 システム障害時のダウンリカバリ

4．メッセージ通番

以上メッセージのリカバリ方式について述べたが、実際は各端末に対する環境のリカバリも同時に行なう必要がある。そのためにシステム状態管理が、端末の状態情報を定期的かつ、端末障害発生時にシステム状態ファイルへ書き出す。リカバリ実行時、そのシステム状態ファイルの内容も使用して環境のリカバリを行なう。

その環境データの一つとして、各端末のメッセージ通番を矛盾のないように管理する通番器があるが、その復元について以下に述べる。

［端末障害］

システム状態ファイルから得た値により入／出力メッセージの通番器を復元する。

［システム障害］

まずシステム状態ファイルにより通番器の復元を行なう。次に、システム状態ファイルに書き込まれなかった部分を補うためにＩＭＪ、ＯＭＪ（各メッセージの通番）、ＣＫＰＪ（全端末のメッセージ通番のテーブル）を利用して通番器の値を調整する。

6．おわりに

本方式では、障害の形態を分析して機能を明確に分けることによりメッセージリカバリを高速に実行し、メッセージ通番を利用することにより、メッセージの順序性、信頼性を保証した。

実際に本方式の採用により、ホットスタンバイ方式において、稼働系から待機系への切り替え時間が、リカバリ処理も含めて約３０秒で実現された。

参考文献

[1]森　他：ﾐﾆｺﾝﾋﾟｭｰﾀでのｼﾞｬｰﾅﾙﾘｶﾊﾞﾘ高速化手法(1)
[2]松井　他：ﾐﾆｺﾝﾋﾟｭｰﾀでのｼﾞｬｰﾅﾙﾘｶﾊﾞﾘ高速化手法(2)
[3]鈴木　他：ﾐﾆｺﾝﾋﾟｭｰﾀでのｼﾞｬｰﾅﾙﾘｶﾊﾞﾘ高速化手法(3)
[4]松井　他：ﾐﾆｺﾝﾋﾟｭｰﾀでのｼﾞｬｰﾅﾙﾘｶﾊﾞﾘ高速化手法(4)
[5]広兼　他：ﾐﾆｺﾝﾋﾟｭｰﾀでのｼﾞｬｰﾅﾙﾘｶﾊﾞﾘ高速化手法(5)
以上情報処理第４１回全国大会(1990)