オペレーティング・システム 43 － 2
（１９８９．６．９）

オンライン高速履歴情報取得制御方式

松沢尚光
（株）日立製作所ソフトウェア工場

銀行・証券業界を中心とした第３次オンラインシステムの構築の動きに対して、システムの性能に重要な要因となる履歴情報の取得方法を構成・制御方式から改善検討した結果について述べる。履歴情報の目的別分割取得の概念を導入し、目的にそった記憶装置に、適切な時点で分割取得する方法を検討した。具体的には外部記憶装置に半導体記憶装置とカートリッジ型磁気テープ装置を適用し、記憶装置の特徴を生かしたことによって、高性能な履歴情報取得機能を実現した。この方法により履歴情報取得性能を向上でき、設備規模の面からも都市銀行の大規模システムに対応可能なものとなった。

On High Speed Method of Logging for Online Systems

Nobumitsu Matsuzawa
Software Works, Hitachi, Ltd.
5030 Totsuka-machi, Totsuka-ku, Yokohama, 244, Japan

This paper deals with a new method of constructing and controlling log information which improves the performance of online systems, and is especially suitable for large-scale banking and stock marketing systems. In this method, log information is divided into several categories and collected on separate storage devices respectively at appropriate time. It improves the performance of logging, and permits feasible size of hardware equipment supporting large-scale banking systems. IC-disk memory units and magnetic tape units of cartrige type are adopeted to realize them.

1　まえがき

近年、銀行・証券業界は第３次オンラインシステムに向けシステムの再構築を図っている。特に都市銀行を中心とした大規模システムでは、システム規模の拡大に対応した信頼性の向上、性能の確保がＤＢ／ＤＣ製品に要求されている。今回、システム全体の性能に対して重要な要因となる履歴情報の取得方式について大規模システムを対象とした新しい方式を考案し履歴情報取得能力の向上を実現した。この方式の構想から実現方式への展開、開発結果及び評価について述べる。

オンラインシステムの履歴情報としては、多様な種別があるが、ここでは特にトランザクションの実行内容を表わすデータベースの更新情報、メッセージの入出力情報などの時系列な順序を保証すべき情報を対象として考える。以下の文中では「ジャーナル」として説明する。

2　従来のジャーナル取得方式の問題点

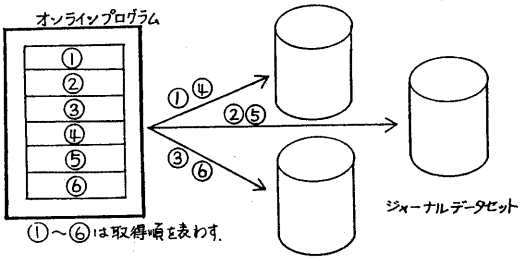
高トラフィックなオンラインシステムではジャーナルの取得時間が性能上のボトルネックとなり易い。この問題点に対する解決策としては、ジャーナルを複数台の外部記憶装置に情報発生順にラウンドロビン方式で取得し負荷を分散する方式がある。この方式を用いても超大規模なシステムでは以下の問題点が発生する事が分った。

(1) ジャーナル取得能力を向上するには、取得媒体（磁気ディスク装置）の台数が膨大になる。

(2) 磁気ディスク装置に取得したジャーナルを磁気テープに保存するためのバックアップ操作が頻繁に発生し、オペレータの負担が大きい。

(3) (2) のオペレータ運用中に不測の障害によりジャーナル取得媒体が満杯となり、オンラインダウンを招く危険性が高い。

上記の問題点が発生する原因は、ジャーナルの



図－１　磁気ディスク装置の負荷分散

持つ複数の目的を一種類の記憶装置にトランザクション実行の特定時点で取得している事に係わっていると考えられる。

3　問題点の解決法

3．1　ジャーナルの目的

ジャーナルの使用される目的はオンライン、オフラインの両方で次のように分けられる。

(1) オンラインダウン時の処理途中トランザクションの回復

トランザクションの回復は、定期的に取得しているチェックポイントダンプを起点としているため、チェックポイントダンプ取得以降の必要な最低限の量が有れば十分である。しかしトランザクションの動作に同期して媒体に出力する必要がある。

(2) データベース障害回復・統計情報解析

データベース障害時のオフライン回復、統計情報解析の目的ではオンライン稼働中の全ての期間のジャーナルが必要である。ジャーナルの取得順で抜けが発生しなければ、媒体への出力タイミングはトランザクションの動作に同期している必要はない。

(3) ユーザ業務処理

オンライン中の業務処理結果をジャーナルとして出力しオフライン業務処理の入力に使用する。この目的では(2) と同様、ジャーナルの取得順で抜けが発生しなければ媒体への

出力タイミングはトランザクションの動作に同期している必要はない。

(4) ジャーナルの保存

トランザクションの実行状況を表わすジャーナルは一定期間保存しておき後で使用することがある。この目的では(2),(3) の目的と重なり、オンライン稼働中の全ての期間のジャーナルが必要である。

上記のようにジャーナルの目的別にその特性を考慮すると、記憶装置の必要な特性が決まってくる。

(1) オンラインダウン時の処理途中トランザクション回復

高速小容量……不揮発性半導体記憶装置が適する

(2) データベース障害回復、統計情報解析
(3) ユーザ業務処理
(4) ジャーナルの保存

中速大容量……磁気テープ装置／磁気ディスク装置が適する

３．２　ジャーナルの構成

３．１の内容を実際のジャーナル構成へ反映することを考える。ジャーナルの使用目的別に適した装置へ、適した取得タイミングで分割取得する

(1) システム回復ジャーナル（ＳＹＪ）

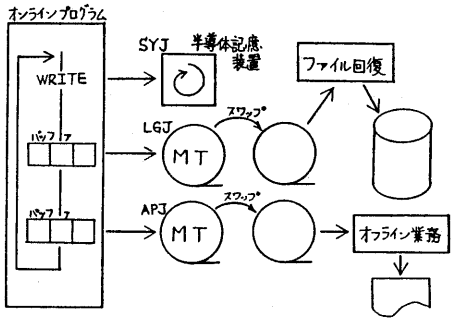
オンライン中のトランザクションの回復に使用するために取得する。半導体記憶装置にラップアラウンド方式で回復に必要な範囲のジャーナルを保存しながら小容量のデータセットに取得する。小容量・高価な半導体記憶装置の適用が可能となる。又、トランザクション非同期出力方式の記録ジャーナル（ＬＧＪ）／アプリケーションジャーナル（ＡＰＪ）のジャーナル情報が失われることがないようにＬＧＪ／ＡＰＪの未出力分を回復できるように管理する。

(2)　記録ジャーナル（ＬＧＪ）

ＬＧＪはデータベースの障害回復プログラムあるいは統計情報解析プログラムの入力用として取得する。トランザクションの実行状態とは同期せずに複数のトランザクション分のジャーナルをブロッキングして出力する。ブロッキングの効果によりＳＹＪより低速の装置でも適用可能である。取得媒体としては磁気テープ装置が適し、オンライン稼働中の全期間のジャーナルを取得し磁気テープで保存する。

(3) アプリケーションジャーナル（ＡＰＪ）

ＡＰＪは業務処理固有の任意の情報を取得しバッチ業務処理の入力に使用する。磁気テープ装置へＬＧＪと同じ方式で出力するか、磁気ディスク装置へ順編成出力する。



図－２　ジャーナルの構成

３．３　性能の検討

性能の検討にあたって次のトランザクションモデルと記憶装置仕様を基にした。

目標トランザクション件数　４０万件／時

ジャーナル出力量

ＳＹＪ：１０ＫＢ／トランザクション
　バッファ長　４ＫＢ×３

ＬＧＪ：　５ＫＢ／トランザクション
　バッファ長　３２ＫＢ

ＡＰＪ：　５ＫＢ／トランザクション
　バッファ長　３２ＫＢ

ハードウェア

磁気ディスク装置

転送速度　　３ＭＢ／秒

回転待ち時間（最大）１６．６ミリ秒

半導体記憶装置

転送速度　　６ＭＢ／秒

アクセス時間先頭ＣＩ　０．７６ミリ秒

２番目以降ＣＩ　０．５２ミリ秒

磁気テープ装置

カートリッジ型

転送速度　　２．５ＭＢ／秒

内部処理時間　　１．０ミリ秒

オープンリール型

転送速度　　１．３９ＭＢ／秒

内部処理時間　　１．９４ミリ秒

磁気ディスク装置と半導体記憶装置のＩ／Ｏ時間を比較すると（ＳＹＪに使用した場合）

磁気ディスク装置　２１．６ミリ秒

半導体記憶装置　３．８ミリ秒

半導体記憶装置では負荷分散を行わなくとも、磁気ディスク装置の６分散と同等の性能を確保できる。磁気ディスク装置を用いた場合、２分散では目標の４０万件／時のトラフィック量でデバイス利用率が１００％を越えてしまう。３分散ではデバイス利用率が８０％程度で危険である。実用上５分散以上は必要と考えると半導体記憶装置の適用効果は大きく、目標のトラフィック量にも対応できる。

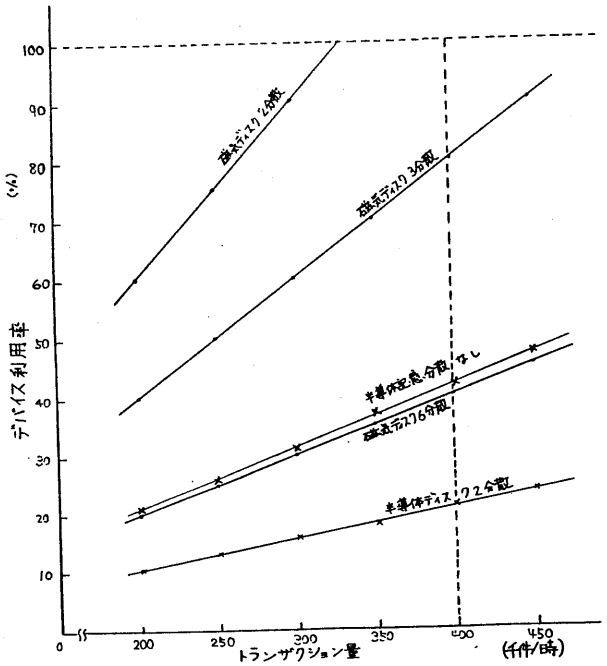
磁気テープ装置のＩ／Ｏ時間をＬＧＪ／ＡＰＪについて計算すると

カートリッジ型　　１３．８ミリ秒

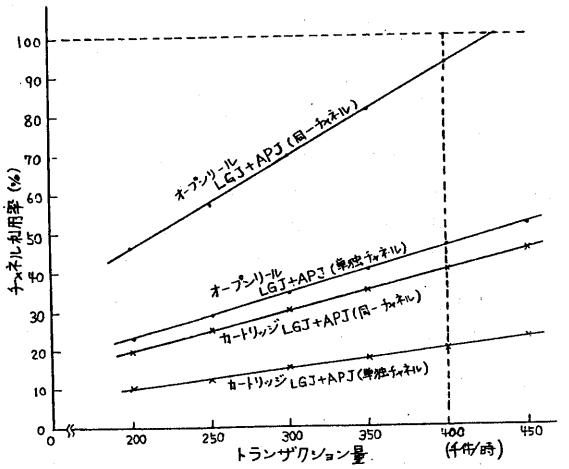
オープンリール型　　２５．０ミリ秒

ＬＧＪ／ＡＰＪを同一チャネルに接続して使用すると目標の４０万件／時のトラフィック量で、オープンリール型ではチャネル利用率が９０％を超えてしまい危険である。チャネルをＬＧＪ／ＡＰＪで使い分けるか、カートリッジ型磁気テープ装置の使用が必須となってくる。

このように３．２で検討したジャーナル構成で目標とするトラフィック量に対応可能なジャーナル取得処理を実現できる見通しがついた。



図－３　ＳＹＪのデバイス利用率



図－４　ＬＧＪ／ＡＰＪのチャネル利用率

３．４　設備規模の検討

磁気ディスク装置に分散取得する従来のジャーナル取得方式と、目的別にジャーナル取得装置を使い分ける新しい方式とで記憶装置の容量を比較し設備規模の大きさについて検討した。

次のモデルシステムによる試算値を示す。

トランザクション件数　　４０万件／時

ジャーナル出力量

全ジャーナル：１０ＫＢ／トランザクション

ＬＧＪ：５ＫＢ／トランザクション

ＡＰＪ：５ＫＢ／トランザクション

ホストオンライン数　　３オンライン

記憶装置容量

磁気ディスク装置

１２００ＭＢ／ボリューム

カートリッジ型磁気テープ装置

約２００ＭＢ／ボリューム

半導体記憶装置

６４ＭＢ／ボリューム

〈従来方式〉　磁気ディスク装置分散使用

ａ　磁気ディスク分散数……５

ｂ　正副二重取得……２

ｃ　データセット数／ボリューム……２

ｄ　世代数（現用・交代用・コピー待ち）……６

ｅ　予備ボリューム数……１

ｆ　オンライン数……３

｛ａ×（ｄ／ｃ）×ｂ＋ｅ｝×ｆ＝９３（ボリューム）

ジャーナルデータセット用装置のみでこのような大規模な構成をとることは現実的には難しい。

〈新方式〉　ＳＹＪ

ａ　半導体記憶装置分散数……４

ｂ　正副二重取得……２

ｃ　データセット数／ボリューム……２

ｄ　世代数（現用）……１

ｅ　予備ボリューム数……２

ｆ　オンライン数……３

｛ａ×（ｄ／ｃ）×ｂ｝×ｆ＋ｅ＝１４（ボリューム）

ＬＧＪ

カートリッジ型磁気テープ装置

ａ　正副二重取得……２

ｂ　世代数（現用・交代用）……４

ｃ　オンライン数……３

ａ×ｂ×ｃ＝２４（デッキ）

ＡＰＪ

カートリッジ型磁気テープ装置

ａ　正副二重取得……２

ｂ　世代数（現用・交代用）……４

ｃ　オンライン数……３

ａ×ｂ×ｃ＝２４（デッキ）

このように従来方式と新方式の装置構成の比較をすると大幅に低減を図ることが可能であり、設置スペースの削減に効果が期待できる。

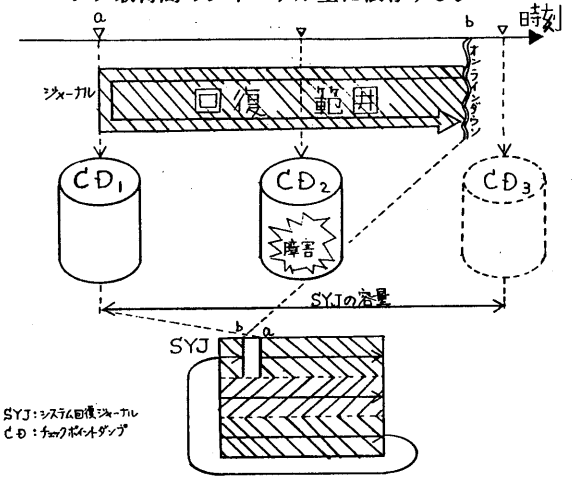
３．５　ジャーナル制御方式

３．２のジャーナル構成にてジャーナルを取得する時の制御方式について述べる。

(1) ＳＹＪのラップアラウンド管理

ＳＹＪはラップアラウンド方式で古いデータに新しいデータを重ね書きしていく。このためにはオンラインダウン時のトランザクション回復に必要なジャーナルを消去することがないよう保証しなければならない。トランザクションの回復にはメモリの情報をチェックポイントダンプを基準点として回復しチェックポイントダンプ取得以降のジャーナルを用いて回復を行う

このためチェックポイントダンプ取得時点前のジャーナルは消去可能としてラップアラウンド管理する。ＳＹＪの容量はチェックポイントダンプ取得間のジャーナル量に依存する。



図－５　ＳＹＪのラップアラウンド管理

(2) カートリッジ型磁気テープ装置の適用

カートリッジ型磁気テープ装置はオープンリール型に比較し装置・媒体の小容積化、操作性の向上を図ることができるのと、磁気テープ制御装置（ＭＴＣ）に大容量のバッファを内蔵しＣＰＵの入出力要求とは非同期に媒体との入出力を行うことによって高速な入出力を実現できる特徴がある。

本来、ジャーナルの取得処理は磁気ディスク／磁気テープ等の外部記憶装置へジャーナルを出力要求すると媒体上へ反映されたことを保証して成立している。しかしカートリッジ型磁気テープ装置はＣＰＵのプログラム処理とは非同期に媒体へ出力される。このためＬＧＪ／ＡＰＪに適用するにあたって、出力タイミングがトランザクションの動作に同期しないＬＧＪ／ＡＰＪの出力方式とは合致するが、次の問題点がある。

ＭＴＣ内のバッファに残留していた分のデータが磁気テープ媒体へ出力されずに消失する可能性がある。

磁気テープ媒体／磁気テープ装置（ＭＴＵ）障害の場合、ＭＴＣ内に残留しているデータをＭＴＣからＣＰＵに読み戻せばデータの消失は防止できる。しかし、ＭＴＣ／チャネル障害の場合、ＭＴＣ内の残留データを読み戻せないので、上記のデータ消失の問題が発生する。

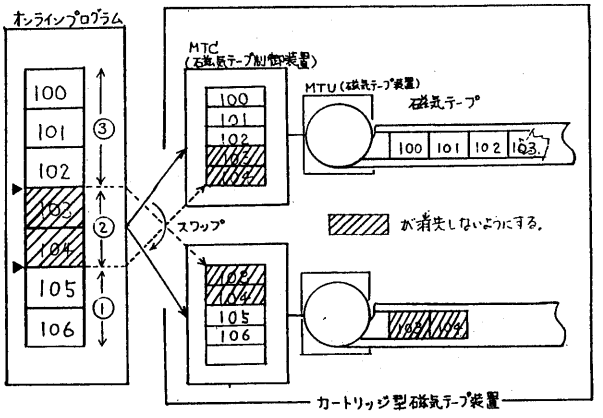
この問題点の対策としてはＬＧＪ／ＡＰＪのバッファを３つの状態に分けて管理することにより実現できる。

① バッファ上にジャーナルデータをバッファリング中であり、ＭＴＣに対して出力要求していない。

② ＭＴＣへ出力要求したが、ＭＴＣのバッファ上に残留して、まだ媒体上へ出力されていない。

③ ＭＴＣへ出力要求しＭＴＣのバッファ上から媒体上へ出力済になっている。

③のバッファは①に変更して新しいデータのブッファリングを続けられるが②のバッファは残留させて置くように管理する。ジャーナル出力



図－６　カートリッジ型磁気テープの管理

中にＭＴＣより障害の報告を受けると、②のバッファをスワップ先に再出力する。これによってジャーナルデータの欠落を防ぐ。この方式の特長は、磁気テープのスワップ時間を短くすることができる点と、ＭＴＣ／チャネルの障害に対しても対応が可能な点である。

４　あとがき

高トラフィックなオンラインシステムに対しジャーナルの使用目的別に、適した記憶装置へ適した取得タイミングで分割取得する考え方は、性能の確保、設備規模の低減を図る有効な方法といえる。（株）日立製作所においては、TMS-4V/SP/J1 (Transaction Management System-4V/System Product/Extended Journal Function 1) の製品名で製品化を行い、昭和６３年秋に本格稼働した都市銀行のシステム他で、性能面・設備面での実績を得ている。

社会のニーズはさらに大規模なオンラインシステムの実現へと拡大しているが、それに対応するための履歴情報の取得方式を本稿の内容を基に発展させていきたいと思う。

参　考　文　献

1) 金居他：履歴情報取得方式，公開特許公報，特願　昭64-21649号（1989).
2) TMS-4V/SP 解説，（株）日立製作所，(1988).
3) TMS-4V/SP 操作，（株）日立製作所，(1988).