リアルタイムシステムにおけるサービス無中断方式の一考察

20－7

白石　正裕、　林　徹、　松沢　寿典、　小南　俊雄

ＮＴＴ情報通信研究所

1　はじめに

オンライントランザクション処理システムにおいて、処理要求に対し即座に応答を返却する高リアルタイム性と、障害が発生した場合もサービスを停止せずにシステムを運用可能な高アベイラビリティ（高可用性）が要望されている。一方、サービスの多様化、高度化によって、サービスを実現させるために扱うべきデータの量も増大しており、複数のプロセッサによる分散システムを構成する場合も多くなってきている。

本稿では、複数のプロセッサで構成したシステムにおいて、ＴＰモニタで複数のサービスを管理することによって、障害が発生した場合に、サービスを中断することなく障害となったプロセッサの業務を実施することが可能な方式について考察する。

2　従来のサービス無中断方式

これまで、高リアルタイム性や高アベイラビリティが要求されるシステムでは、ホットスタンバイ方式のように、障害発生時には、現用系で実施していたサービスを予備系に引き継ぎ、サービスを継続させる方法があった。(1) 一方、複数のプロセッサで構成する分散システムの場合には、データベースやオブジェクトのレプリケーション（複製）の機能を用いて、障害が起きた場合に、システムを停止せずに運転可能な方式も紹介されている。(2) しかしながら、従来の方式の場合、基本的に予備系ではサービスを実行していないため、プロセッサ資源の無駄である。また、二重障害（予備系で運用中の障害）が発生した場合、システムを停止せざるを得なくなり、重要な更新情報が欠落することになりかねない。

3　リアルタイムシステムにおけるサービス無中断方式

3．1　対象となるシステム条件

本稿で対象とするシステムの条件として、以下に明確にする。

（1）システム構成（図１参照）

高リアルタイム性、及び高アベイラビリティが要求されるシステム構成として、通信を介して接続され、地域に分散した複数のプロセッサからなる疎結合のシステム構成を考える。

各プロセッサには、リアルタイム性を考慮して、アクセス高速化のためにメモリ上にサービスに対応したデータベースを所有し、サービスのトランザクション管理を行うトランザクション処理モニタ（ＴＰモニタ）を配備し、サービスを実行する。

ＴＰモニタにおいては、主にトランザクションの開始／終了、及び障害時のリカバリ処理に関する管理等を実施する。

A Study of Service Nonstop Method in Real-time Computer System

Masahiro Shiraishi,Touru Hayashi,Hisanori Matuzawa, Toshio Kominami

NTT Information and Communication Systems Laboratories

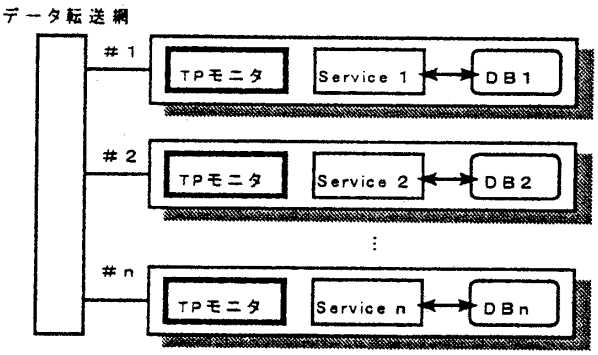

図１：システム構成

（2）高可用性

障害が発生した場合、短時間にバックアップ側で運用可能であり、またサービスを中断させずに運用可能である。

（3）トランザクションの保証性

システムで処理するトランザクションは必ず保証される必要がある。障害によって、トランザクション処理による実行結果が欠落してはならない。

以下に上記システム条件を満足する無中断方式について述べる。

3．2　サービス無中断方式

システム稼働中に障害が発生した場合において、サービス無中断でシステムを運用する方式の概要を以下に示す。

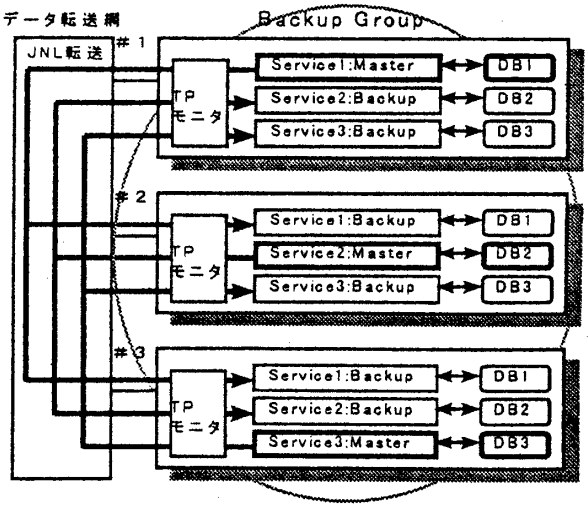

図２：バックアップ構成（3プロセッサ構成の場合）

（1）バックアップグループの形成（図２参照）

複数のプロセッサ間において、バックアップグループを形成する。バックアップグループ内において、あるプロセッサが障害により停止した場合は、バックアップグループで定義された他プロセッサにて障害プロセッサの業務を継続して実行する。２プロセッサ以上からなるバックアップグループを形成するこ

とによって、二重障害（バックアップ側での障害）が発生した場合も、バックアップグループの他プロセッサによって業務を続行することが可能となる。

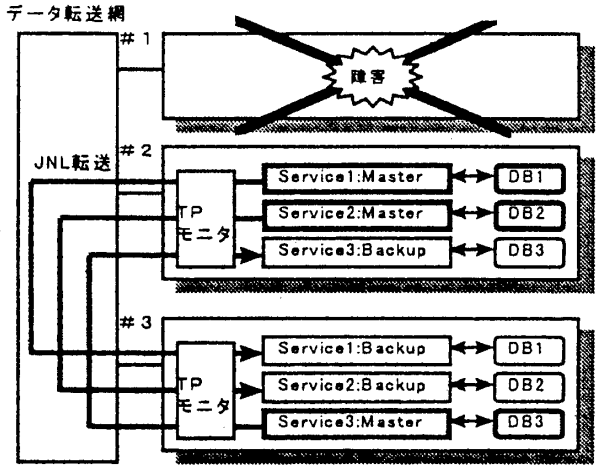

図3:障害時の切り替え方式(3プロセッサ構成の場合)

（2）相互バックアップの実現（図3参照）

全プロセッサにおいて、バックアップグループ単位で相互にバックアップを実施する。つまり、あるプロセッサにおいて、自プロセッサで実施しているサービス（マスタサービス）とは別に、バックアップグループで定義された他プロセッサで実施しているサービスのバックアップ（バックアップサービス）を保持する。マスタ／バックアップサービスのトランザクション管理はＴＰモニタにて実施する。これによって、あるプロセッサにおいて、障害が発生した場合にでも、他プロセッサにて、自サービスを継続させながら、障害側プロセッサの処理を継続して実行することが可能となる。

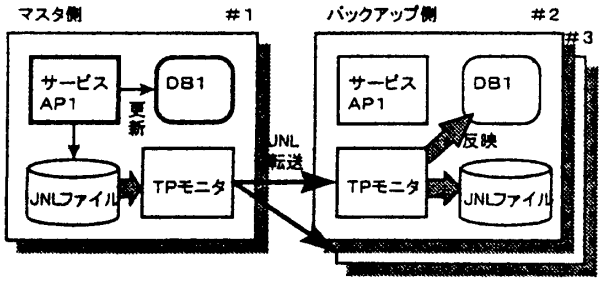

図4:ジャーナル反映方式

（3）各サービスにおけるトランザクションの保証（図4参照）

各プロセッサで保持しているマスタ／バックアップサービスのトランザクションの保証は、ＴＰモニタが管理する。あるプロセッサ上のＴＰモニタでは、マスタサービスのトランザクションの保証とともに、バックアップグループで定義された複数のプロセッサに対して、マスタサービスの実行によって発出したデータ更新履歴情報（以下、ジャーナル）を転送する。バックアップ側プロセッサ上のＴＰモニタでは、転送されたジャーナルをもとに、メモリ上のデータベースに反映する。上記処理は、バックアップグループで定義された複数のプロセッサ間において、相互に実施する。ジャーナル転送契機は、定周期もしくは、一定ジャーナル量の蓄積毎に実施する。これによって、バックアップグループ内のメモリ上データベースの更新状態を最新とし、障害時にバックアップ側プロセッサへの高速な切り替えが可能となる。

4　本方式の評価と考察

4．1　評価

本方式を実現する上での評価パラメータとして、以下の項目について評価した。

（1）プロセッサ資源の使用効率

複数のプロセッサにて相互にバックアップを実現するため、バックアップのために予備プロセッサを用意する必要がない。また、マスタサービスを実行しているプロセッサ上で、バックアップサービスを実行するため、プロセッサ資源の無駄がない。

（2）通常時のオーバヘッド

定周期毎にバックアップ側プロセッサに対して、ジャーナルを転送しているため、転送するジャーナル量が多くなり、通信路の負荷が増大する。また、１プロセッサ内において、複数のバックアップサービス分のメモリを確保する必要があること、及び複数のバックアップ処理を実行することによって、メモリの圧迫、ＣＰＵ負荷が増大する。

（3）ＡＰ設計の容易さ

各プロセッサ上で、ＴＰモニタにてバックアップサービスを実行するため、ＡＰ設計者はＡＰを設計する上でバックアップサービスを意識して作成する必要がない。ＡＰの設計には影響なく、バックアップを行うことが可能となる。

4．2　考察

上記評価により、通常時のオーバヘッドが問題となる。本方式を実現する上で、データ転送路の設計に関して、我々が設計の対象としているシステムをモデルとして、考察を行った。

（1）設定条件

１プロセッサ上のマスタサービスに送信される最繁時のトランザクション量を 150Tr/s、１トランザクションあたりに発出するジャーナル量を 512byte とする。

（2）データ通信路の設計

サービスの実行によって発出するジャーナル量は、76.8kbyte/s＝614.4kbps である。バックアッププロセッサ間で相互にバックアップを行う場合、614.4k＊2≒1.2Mbps のデータ転送速度が必要となる。我々のシステムでは、1.5Mbps の専用回線を複数回線用いて実現する方向である。

5　おわりに

複数のプロセッサで構成されたシステムにおいて、サービス無中断にてサービスを継続することが可能な方式について述べた。

参考文献

[1] Ｊ.グレイ他、渡辺訳：“ＯＬＴＰシステム”、マグロウヒル出版

[2] 斉藤他：“データ分散化とオブジェクト再構築に基づく分散処理システムの高信頼化方式“、情報処理学会論文誌（１９９５）