

クラスタ/並列マシン上でのOLTPシステムの構築方式の検討

5N-3 五十嵐 明夫\* 岩崎 元昭\* 宇佐見 憲司郎\*\* 山本 展之\*\*\*

(株)日立製作所 ソフトウェア開発本部\* (株)日立西部ソフトウェア\*\* (株)日立製作所 システム開発研究所\*\*\*

1 はじめに

UNIXを搭載するRISC WS/サーバの性能向上は主としてマシンサイクル時間の短縮に負うところが大きかった。しかし、1つのシステムが必要とする処理能力は、単一プロセッサの能力では補えないところまで大規模化してきており、今後の性能競争ではマルチプロセッサ技術が重要になると考えられる。このような技術的背景から、高性能RISC WS/サーバをLANや専用高速通進路等のH/Wを介して疎結合したサーバ群(クラスタ/並列マシン)が性能を追及する点から重要になると考えられる。

そこで、弊社の分散トランザクションモニタ OpenTP1を使いクラスタ/並列マシン上に、基幹系業務システムにも耐えうる信頼性を確保したOLTPシステムの形態を提案する。

2 クラスタ/並列マシン上のOpenTP1システムの構成

OpenTP1は弊社のメインフレーム上のTPモニタXDMの技術を取り入れており、信頼性の高いTPモニタを実現している。これをクラスタ/並列マシン上に搭載する際の基本方針として次の2つが考えられる。

- (1) マイクロカーネル的な考え方を採用し  
OpenTP1を構成するシステムサーバ群を機能的にサブセットに分割し、クラスタ/並列マシンを構成する各ノードにそのサブセットのみを搭載する。
- (2) OpenTP1を構成するシステムサーバ群は、そのままクラスタ/並列マシンを構成する各ノードに搭載し、クラスタ/並列向け機能は、各ノード上のOpenTP1間の連携強化機能として位置付ける。

(1)はクラスタ/並列マシンを構成する各サーバの回復リソースを特定サーバが保持することになり一点障害の危険性が増し、高信頼性が最重要要件となる基幹系業務システムには適用し難い。そこで、(2)を基本方針としてシステムを構築した。概要を図1に示す。

“OLTP system on cluster/parallel machine”  
Akio Igarashi, Motoaki Iwasaki, Kenshirou Usami,  
Nobuyuki Yamamoto  
\*Software Development Center, Hitachi, Ltd  
\*\*Hitachi Seibu Software, Ltd  
\*\*\*Systems Development Laboratory, Hitachi, Ltd

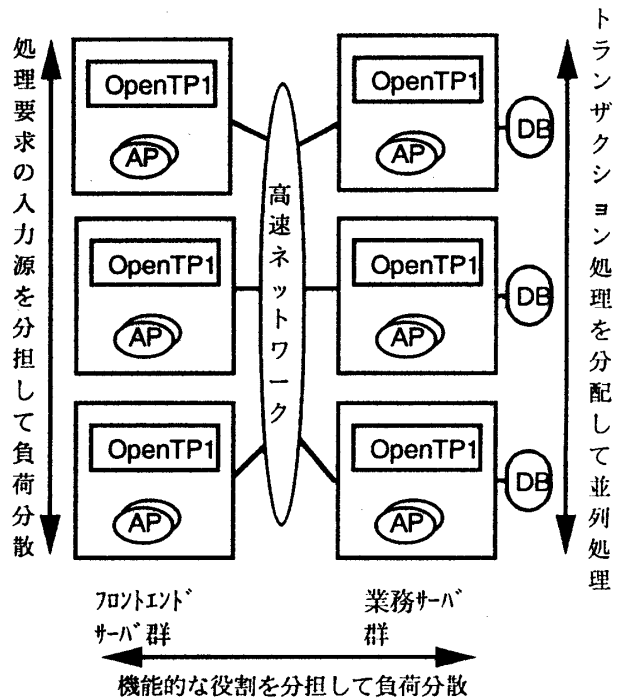


図1 システム構成

フロントエンドサーバ群は、クライアントからのサービス要求(メインフレームからの回線経由の要求、WS/PCからの要求等)を入力源毎に分担して処理する。

フロントエンドサーバ群で受け付けた要求は、クラスタ/並列システム内の高速ネットワーク(バス)を通して業務サーバ群に渡される。

各サーバは、OpenTP1の種々の回復機能をそのまま保持しており、特定ノードの障害がクラスタ/並列マシン全体のシステム障害となることはない。

3 OpenTP1の連携強化機能

クラスタ/並列マシンを効果的に利用するために、OpenTP1では次の機能を提供する。

- (1) スケーラビリティの向上  
負荷に応じてプロセス数を自動的に増減させる。更に、特定ノードにサービス要求が集中しないようノード間の負荷分散スケジューリング機能を実現する。
- (2) 高信頼性  
ノード障害時、系切替えにより処理を続行する。更に、Oracle等の他社DBとの連動切り替えを実現する。

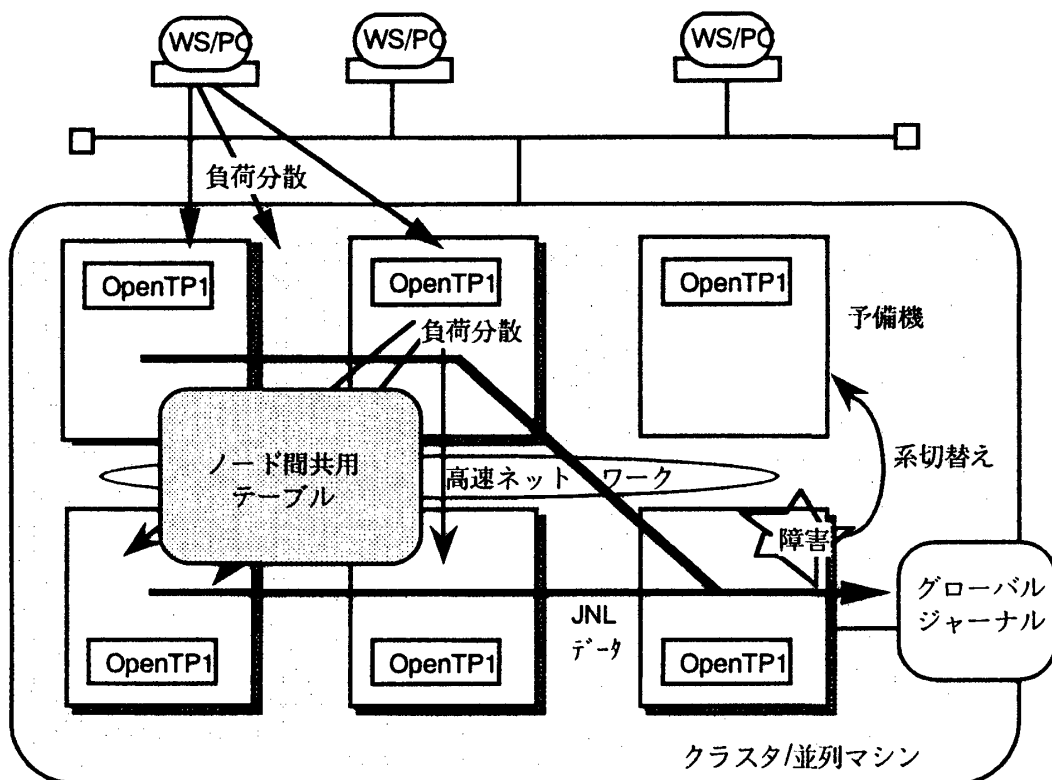


図2 OpenTP1の連携強化機能

(3)高性能

高信頼UDP/IP通信を前提として、高速なRPC機能を実現する。

(4)一元運用

各ノードで出力されるファイル更新ジャーナルやユーザジャーナルをリアルタイムでグローバルジャーナルファイルに収集しジャーナル運用一元化を実現する。

(5)業務システム構築支援

複数ノード間でデータを共用できるメモリ上のテーブルを実現する。(1つのノードにおける更新処理を全てのノードに伝播させる)

以上の機能を、図2及び表1にまとめる。

表1 OpenTP1の連携強化機能

要件	OpenTP1の対応
スケラビリティの向上	自動負荷分散スケジューリング
高信頼性	系切り替え
高性能	高速RPC
一元運用	グローバルジャーナル
業務システム構築支援	ノード間共用テーブル

4 おわりに

本システムでは、クラスタ/並列マシン上に基幹系業務システムにも耐えうる高信頼なOLTPシステムを構築することを可能にした。

今後は、数十/数百ノードからなるクラスタ/並列システムのシステムチューニングや性能チューニングの手法を確立していく必要があると考える。

5 参考文献

[1] Jim Gray Andreus Reuter, "TRANSACTION PROCESSING", Morgan Kaufmann Publishers, 1993  
 [2] Ira Krepchin, "UNIX Server Power Grows In Clusters", DATAMATION, AUGUST 15, 1993, pp.57-59