

1.はじめに

現在、データベースシステムは、ダウンサイジング、情報技術の向上により、ハードディスクの大容量化となりましたが、保存したデータをいかに迅速かつ、確実に引き出すかが問題である。そして、少数でデータベースを利用している時には問題が起これなくとも、複数のユーザーが同時に利用した時、パフォーマンスが低下し、システムがダウンするという問題が発生する。

この論文では、分散データベース環境で上記の問題が起こる確率を低くし、可用性、信頼性が高いシステムを提案する。そのシステムとは、分散しているデータベースサーバーにネットワークの状態を把握してアクセスするシステムを構築する。複数のユーザーから検索情報を受信して、どのデータベースサーバーへのネットワーク帯域が空いているか把握し、そのデータベースサーバーにアクセスして結果を迅速に返すのである。かつ、データベースの変更は、複数のデータベースに信頼性の高いアクセスを提供する。

2.システムの概要

ここでは、このシステムの概要を説明する。このシステムは、分散しているデータベースサーバーにマルチアクセスをして、ユーザーに迅速かつ、信頼性の高いアクセスを実現する。図1に示すのは、システムの構成図である。

マルチサーバーとは、クライアントの応答、データの受信や送信、データベースのアクセス、ネットワークの状態監視など、このシステムの総合的な処理をしている。

クライアントは、マルチサーバーにアクセスして、情報を取得するデータベースクライアントである。

データベース1、データベース2、データベース3の構成は、共有データ型、分散データ型どちらでも対応可能である。

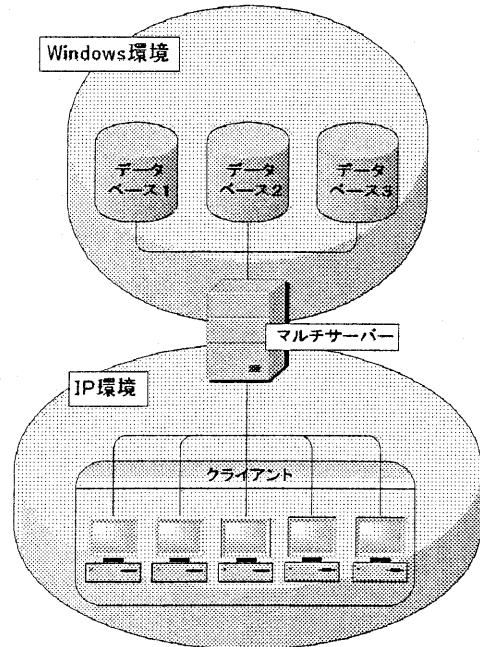


図1.システム構成図

図1で示している Windows 環境と表示されている範囲は、NetBEUI のネットワーク環境である。IP 環境と表示されている範囲は、インターネットで特に使用されているグローバルな IP アドレスのネットワーク環境なので、このシステムをインターネット上で利用する事も可能である。

3.システムの特徴

- ・実際には複数のデータベースサーバーで構成しているシステムだが、クライアントからはマルチサーバーにアクセスする単一のサーバーであるかのように処理を行っているので、クラスタリング環境を実現している。

- ・データベースは、共有データ型、分散データ型、どちらでも対応可能である。共有データ型の場合は、一台のデータベースが何らかの原因でダウンしても、残りのサーバーがカバーできるので可用性が高い。分散データ型の場合は、あるデータを一つのサーバーに保持するだけなので、共有データ型よりもディスク容量を使わないのである

- ・マルチサーバーとクライアントの間は、IP アドレスとポートを使用したソケット通信しているので、このシステムは、小規模な LAN だけではなくグローバルな TCP/IP 環境で動作可能である。ここでのソケットは WinSock2.0 を使用している。

- ・マルチサーバーはデータベースにアクセスする時、並列処理を行うマルチスレッド処理を行っている。スレッドとは、プロセスのコードやデータセグメントをプロセス中のほかのスレッドと共有するが、レジスタ値や、スタック、実行中の命令アドレスはスレッドごと固有である。そのため、それぞれのスレッドに CPU 時間が与えられるので、プロセス中のすべてのスレッドを同時に実行することができる。

- ・データベースにアクセスする時は、マイクロソフトで提供されている ADO(ActiveX Data Objects)2.5 を使用する。データベースの種類は、ADO に対応しているデータベースならこのシステムに適應できるので、Microsoft SQL server, Microsoft Access 2000, Oracle など異種データベースが可能である。ここでは Microsoft SQL Server7.0 を使用する。

4. クラスタリング環境

クラスタとは、単一のシステムとして動作する、複数の独立したシステムのグループのことである。このシステム環境では、クライアントは、マルチサーバーに対してソケット通信でデータの送受信を行っている。そこでマルチサーバーがクライアントからデータを受信したら、登録してある1台もしくは複数のデータベースにアクセスしても、クライアントからは見えない処理なのである。これは、サーバー環境が変化してもクライアント環境は今までのまま使用できるので拡張性に優れている。かつ、複数のデータベースサーバーが同時動作しているので、可用性と信頼性に高いのである。

5. 共有データ型

共有データ型とは、各データベースに同じデータを保持する構成の事である。

- ・クライアントがアクセス情報をマルチサーバーに依頼した時、その情報が選択条件なら、マルチサーバーと各データベースサーバーの間のネットワーク帯域を調べ、一番帯域が空いているデータベースサーバーにアクセスする。そして、結果をクライアントに返すので、データベースアクセス速度が速いのである。

- ・クライアントがアクセス情報をマルチサーバーに依頼した時、それが入力、更新、削除の条件なら、登録してあるすべてのデータベースにアクセスする。そこで信頼性を高めるために各データベースにトランザクションを作成し、2 フェーズコミットでアクセスしているので、信頼性の高いマルチアクセスにしている。

6. 分散データ型

分散データ型とは、各データベースに異なるデータを保持する構成の事である。

- ・クライアントがマルチサーバーに選択条件のアクセスを依頼した時、各データベースは異なるデータを保持しているのですべてデータベースにアクセスして結果を一つにして返すのである。

- ・クライアントがマルチサーバーに入力条件のアクセスを依頼した時、一番帯域幅が空いているデータベースにアクセスする。

- ・クライアントがマルチサーバーに更新、削除条件のアクセスを依頼した時、各データベースにアクセスして結果を返すのである。

7. おわりに

分散データベース環境でのマルチアクセスについて述べた。今後は、ネットワークの帯域状況だけではなく、様々な状況を考慮したデータベースアクセスを検討する。

参考文献

Microsoft MSDN ライブラリ 2000 年 4 月