

Disk DB, Memory DB and, What on Next?

Altibase Technical White Paper

2007/10/01

著作権

このドキュメントに記載されている情報は、このドキュメントの発行時点におけるアルティベース、及びシアンズ・アールの見解を反映したものです。アルティベース、及びシアンズ・アールは市場の変化に対応する必要があるため、このドキュメントの内容に関する責任を問わないものとします。また、発行日以降に発表される情報の正確性を保証できません。

このホワイト ペーパーに記載された内容は情報の提供のみを目的としており、明示、黙示または法律の規定にかかわらず、これらの情報についてアルティベース、及びシアンズ・アールはいかなる責任も負わないものとします。

このソフトウェアおよびマニュアルは、本製品の使用許諾契約書のもとでのみ使用することができます。このソフトウェアおよびマニュアルのいかなる部分も、アルティベース、及びシアンズ・アールの書面による許諾を受けることなく、その目的を問わず、どのような形態であっても、複製または譲渡することは禁じられています。ここでいう形態とは、複写や記録など、電子的な、または物理的なすべての手段を含みます。

アルティベース、及びシアンズ・アールは、このマニュアルに記載されている内容に関し、特許、特許申請、商標、著作権、またはその他の無体財産権を有する場合があります。このマニュアルはこれらの特許、商標、著作権、またはその他の無体財産権に関する権利をお客様に許諾するものではありません。

特に記載していない場合、例として登場する企業、組織、製品、ドメイン名、電子メール アドレス、ロゴ、人物、場所、およびイベントはすべて架空のものです。実在する企業、組織、製品、ドメイン名、電子メール アドレス、ロゴ、人物、場所、またはイベントとは一切関係ありません。

© 2007 ALTIBASE Corporation. All rights reserved.

著作權	2
Introduction.....	4
Change in Environment of DB Application.....	4
Limit of Disk-Based DB	5
Limit of Memory-Based DB.....	5
Problems of Operating both DRDB and MMDB.....	6
A New Alternative	7
New demand of Customers.....	7
Consideration of Data Weight.....	7
Suitability of DBMS in various operating environments.....	8
Altibase : Enterprise Main Memory DBMS	8
Hybrid DBMS.....	8
Altibase Concepts	8
System Architecture.....	9
Altibase Features.....	10
Concurrent Accessibility to Memory Table and Disk Table	11
DML (Data Manipulation Language).....	12
Join.....	12
Sub-query.....	12
Retrieval by View	12
Three Operating Models	13
MMDB Model	13
DRDB Model.....	13
Hybrid DB Model.....	13
Complete DB Recovery.....	14
DB Replication	14
Upgrade without Shutdown	15
High Performance with Low Investment Cost.....	15
Investment Cost	15
Performance	16
Conclusion	17

Disk DB, Memory DB and, What on Next?

リレーショナルDBMS(データベース管理システム) は1980年代に紹介されてから20年経過しました。これまで様々な研究・開発を経て、IT技術の中でも中核技術として発展してきました。DRDBMS (Disk Resident Database Management System: ディスク常駐型のデータベース管理システム) は基本的なDBMS、データウェアハウス、データマイニングなど様々な領域で利用されています。しかし、これらDRDBMSはハードディスク上に全てのデータを配置するという前提を置いた構造を持っているため、Disk I/Oがネックとなるパフォーマンス上の制限があります。従って、高いパフォーマンスが求められる分野への適用についてはいくつか困難な問題が出てきます。この技術文書は、これらパフォーマンスの問題を克服したもう一つのDBMSについて紹介するものです。

Introduction

In mass storage DB, there is data requiring high performance

Change in Environment of DB Application

ここ最近のインターネット、モバイル通信などの分野を俯瞰すると、より多くのサービスをより多数のユーザへ提供する事例が増えつつあります。データベースアプリケーションの特徴としては多数のユーザからの同時サービス要求を即座に処理することだといえます。例えば、人気のあるインターネットポータルサイトやインターネットゲームサイトなどは数十万～数百万規模の会員を持ち、ユーザ認証処理だけを見ても、秒間数百件以上のトランザクションを処理できるDBMSが求められます。モバイル通信分野のHLR (Home Location Register) システムなどにおいては、一秒間に約2500コール以上の処理能力が求められます。これらはDBMSに到達するトランザクション数に換算すると約15,000トランザクション/秒以上の性能要件となります。従って、DBMSアプリケーションに求められる要件は、高速に処理して高速に応答を返すだけでなく、同時に多数のリクエストを処理する能力も求められます。

一方、障害対策面においても要求レベルは日々高まっている状況だといえます。例えばインターネットトレードサイト、リアルタイム課金システムなどでは24時間サービスが求められます。リアルタイム課金システムなどにおいては、なんらかの障害により仮に一秒でも処理が停止すると莫大な金額損失を被る事態に直結します。もちろんこれらのシステムは基本的な要件として短時間で処理を完了させ応答を返すことが求められます。実際、携帯電話におけるリアルタイム課金システム分野においては、一秒間に約 4,000 CD R (Call Data Record) の処理能力が求められます。従って、このようなシステムのDBMSとしては迅速な障害回復の仕組み、及びトランザクションの高速な処理と高速なレスポンスタイムを保証することが必須となります。

Limit of Disk-Based DB

このような状況の中、DBMSに求められるものは、高速レスポンス性能だけでなく、拡張性なども同時に求められるようになってきています。ディスクベースのDBMSは多くの機能、多くの運用支援ツールを持ち、あらゆるニーズに対応できるものとして一般的になっていますが、ディスクに全てのデータを保存する方式を採っているためパフォーマンス改善上避けられない構造上の問題を内包しており、レスポンスタイムがこれ以上引き出せない状況を抱えています。従って、本来の性能改善に限界があることからDBクラスタなどでパフォーマンス問題に対応している面があります。但し、クラスタでは全体の処理能力は上がってもサーバ台数が増加してしまうという問題、及び高速レスポンスの実現という点においては何も解決されません。

信頼性の面においては、DBクラスタにより要件を満たせる場面がありますが、クラスタを構成する場合の条件としてクラスタサーバ間の通信、複数サーバ間での共有データの変更管理など様々な要素が絡み合い、高性能処理が求められる分野への適用では十分な性能が引き出せない面があります。

Limit of Memory-Based DB

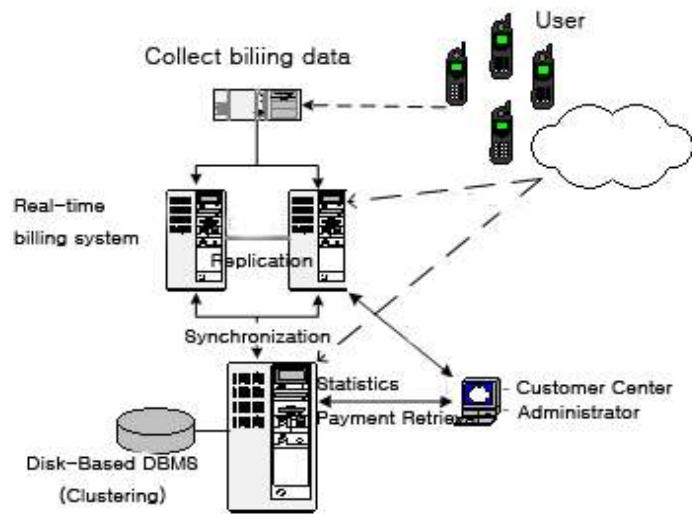
ここ数年でMMDBMS (Main Memory Database Management System:メインメモリデータベース管理システム)の利用が増加してきています。メモリDBは前述したデータベースの利用領域に加えて、より高速なパフォーマンスが要求される分野に浸透してきています。例えば、DLR (Data Location Register)、Intelligent Network Service(SMP/SCP,SMS,...)、インターネット統合認証システム、現物/先物金融商品のリアルタイム分析、ゲームポータ

ルサイトなどです。これらの領域では24時間のサービスが求められ、DBを二重化して冗長構成を持たせることで対応しています。物理メモリは複数サーバ間で共有できないため、一般的にメモリDBではデータベースレプリケーション機能をサポートしています。これはディスクベースのDBクラスタよりも高速な処理が可能です。しかし、メモリDBにも課題があります。コスト的な問題で大容量データをメモリ上に載せられないという点です。メモリDBでは全てのデータを搭載できるだけの物理メモリが必要になります。

Problems of Operating both DRDB and MMDB

これまで述べてきたとおり、ディスクDB と メモリDB はそれぞれ一長一短があります。ディスクDBはパフォーマンスの制限があり、メモリDBはデータサイズの制限があります。これらの制限を解決するために両方のDBMSを使って構成するシステム事例もあります。例えば、携帯電話の課金システムでは直近3ヶ月分のデータをメモリDBに配置し、メモリDB上のデータを含めて全てのデータをディスクDBに配置する方式です。(図 1)

Would you take a complicated structure to get speed of couple thousand TPS?



<図 1> Operating both DRDB and MMDB

このような構成をとる理由は、携帯のユーザだけでなく携帯電話サービス会社側も直近3ヶ月のデータを頻繁にアクセスするためです。また、携帯ユーザが電話をかける際、そのコールに関係する課金情報をリアルタイムに処理し、月次の請求データへ反映する処理があることも挙げられます。これらの情報は全てメモリDB上に配置されます。携帯の利用者はいつでも自分が利用した課金情報をリアルタイムに取得することが可能になります。

メモリDB上に配置された3ヶ月分のデータを含む全てのデータはディスクDB上で管理されます。これらのデータを全てメモリDBに配置するにはコスト的に無理があります。また課金統計情報や顧客管理情報などについては特に高速性が求められるメモリDB上に配置する必要もありません。

このメモリDBとディスクDBの混合構成では高いパフォーマンスと大容量データの管理の両面のメリットを享受できますが、以下の問題を抱えています。

- 二つのDBMSを購入する必要がある
- アプリケーションが二つのDBMSを意識する必要がある
- メモリDBとディスクDB間でデータの同期を管理する必要がある
- 二つのDBMSを運用、保守していくのはコストが嵩む

A New Alternative

Now, there are various classes even in data. Altibase 4 is a new DBMS which considers data weight.

New demand of Customers

システムは通常、全ての必要機能を一つの情報システムで処理できるように計画します。また、システム構築後の運用と保守フェーズを考慮し、なるべくシンプルな構成にします。データベースについても例外ではありません。従って、システム的设计者、開発者、運用管理者は共に一つの共通した環境上でシンプルな構成を採用する方向に動きます。前述の構成は、高速パフォーマンスと大容量データ管理の両面を解決できる反面、いくつかの問題があります。このような状況でDB開発者からはメリットを残したまま二つの異なるDBMSをシームレスに利用できる新たなDBMSが求められるようになりました。

Consideration of Data Weight

既存のディスクDBは全てのデータをディスク上に配置して管理します。言い換えれば全てのデータは頻繁に参照、更新されるかを問わず全て同一属性を持ったデータとして扱われます。ディスクDBとメモリDBを用いたアーキテクチャでは頻繁に参照、更新されるデータはメモリDBに配置されます。これはお互いのDBMSの特徴の違いが存在する故にデータの扱い方も異なってくるということを意味しています。もし、これら二つの違いを同一DBMS

上で吸収し管理できれば、コスト面、管理面、性能面などあらゆる面でメリットがあるでしょう。

Suitability of DBMS in various operating environments

全てのデータベースが高速性と大容量データ管理能力を求められるわけではありません。データベースはデータのサイズ及び性能要求面から次の3つのパターンに分けられます。

一つ目のパターンは、HLR、DLRなど通信分野で必要とされるDBです。これらDBのサイズは数十から数百MBの単位です。この分野で求められる最優先課題は高速データ処理性能です。秒間数千件の処理能力が求められますから通常のディスクDBでは達成が困難な数字となります。達成するにしてもそれ相応のコスト、構成が必要となります。

二つ目のパターンは、ディスクDBがグループウェア、会計システムなど一般的な社内システムの中で利用されるケースです。これらのシステムでDBに求められるのは、高いパフォーマンスではなく、メモリDBにとってはサイズが大きすぎる大容量データを低コストで管理する能力です。

三つ目のパターンは、前述したリアルタイム課金システムのようにDBサイズも大きく、さらに高速処理性能が求められるシステム分野です。一般的に、このようなシステムで高速処理性能が求められるのは、全データベースの内の一部であり全てではありません。

以上3つのパターンを挙げましたが、これらほとんどのケースに適応可能な新しいDBMSがEnterprise メモリDBのAltibaseです。

Altibase : Enterprise Main Memory DBMS

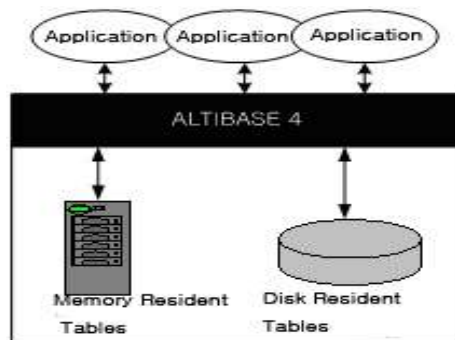
Hybrid DBMS

Altibase Concepts

Manage high performance data in memory tables and other data in disk tables..

ハイブリッド DBMS とは、物理メモリとハードディスクなど異なる記憶装置上にその特徴に適応した配置方法でデータを格納できる一つの新しいデータベース管理システムです。Altibaseは頻繁にアクセスされ高速に処理する必要のあるデータをメモリ上に配置し、それ以外のデータはディスク上に配置することができる新しいコンセプトを実現したDBMSです。

次図にAltibaseの概念図を示します。(図 2)



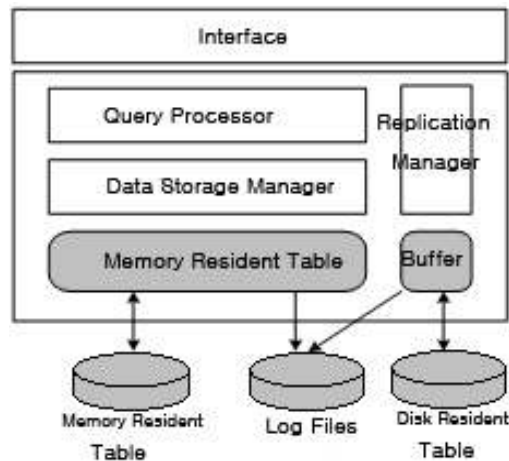
<図 2> Data Storage Architecture of Altibase

System Architecture

Altibaseのシステムアーキテクチャは、大きく3つのパートで構成されます。DBエンジン、インタフェース、ディスク上のファイルの3つです。DBエンジンはさらに三つの構成要素で成り立ちます。メモリ上に格納するテーブル、ディスク上に格納するテーブルを管理するData Storage Manager、SQL文を処理するQuery Processor、DBの可用性を確保するReplication managerです。特にData Storage Managerは、データをメモリ上に配置した後のメモリテーブルデータの管理、またディスクテーブルで使用されるメモリバッファ上に載せられているデータの管理などにおいて、ディスクI/Oをデータの信頼性を損なわずにできるだけ削減して最適化し、さらに高速にデータ処理することが求められます。

ディスク上のファイルは大きく3つに分類されます。メモリ配置テーブルをディスク上に保存しておくファイル、ディスク配置テーブルをディスク上に保存しておくファイル、そして全てのトランザクションログを保持するログファイルです。メモリ配置テーブルについては、Altibaseのエンジンが開始した時、ディスク上のファイルを全て読み込み、メモリ上にデータを展開します。ディスク配置テーブルについては、データがアクセスされた時にページ単位でバッファ領域にデータが読み込まれます。

インタフェースについては、C/C++/JAVA/ODBCなど主要なものをほぼ全てサポートしています。



<図 3> System Architecture of Altibase

Altibase Features

Build a Database with high performance through convenience of development and simplicity of administration

一つのSQL文でメモリテーブルとディスクテーブルにアクセスすることが可能です。例えば、メモリテーブル上の顧客テーブルとディスクテーブル上のトランザクションテーブルを顧客IDで結合する二つのテーブル間のJoinオペレーションも通常のテーブルのJOINオペレーションと同様に処理することができます。

多量のレコードを格納したテーブル (大容量テーブル)はディスクテーブル上に、直近の頻繁にアクセスされるテーブルはメモリテーブルに配置して、参照 (SELECT) 性能を改善することができます。例えば、直近一ヶ月のデータを持ったT(m)という名前のメモリテーブルとトランザクションデータを持ったT(d)という名前のディスクテーブルがあり、これら二つのテーブルT(m)とT(d)を参照するT(v)というビューを作成し、T(v)の中では、最新一ヶ月分のトランザクション情報を参照して性能を向上させる事ができます。

AltibaseのメモリテーブルとOracleのSGAに固定されたテーブルとの違いは次の通りです。まずSGAに固定されたテーブルはメモリバッファからディスクにスワップされる事はありませんが、データアクセス方法に違いがあります。Oracleでは通常のテーブルとSGAに固定されたテーブルの両方のアクセス方法はほぼ同じですが、Altibaseのメモリテーブルは、メモリ上に展開されるデータの格納構造とデータ検索アルゴリズムがメモリ上のデータアクセスであることを前提に最適化されているため、

参照性能が良くなっています。

Altibaseはメモリテーブルのために最適化されたインデックス構造とディスクテーブルのために最適化されたインデックス構造を各々サポートしているため、それぞれのテーブルで特性に応じた最適なパフォーマンスを発揮します。

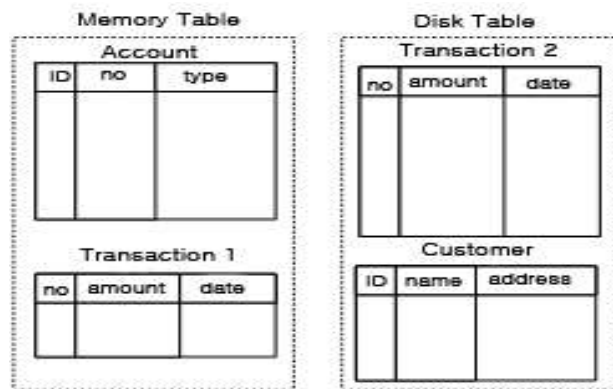
Altibaseではクエリーオプティマイザの強化も行っています。以前のバージョンのAltibase 3 ではhint文でクエリーを最適化する必要がありましたが、現在のバージョンでは、クエリーのhint文なしに最適化するように改善されています。

Altibaseは、メモリテーブルとディスクテーブル間のレプリケーションをサポートしています。レプリケーション定義が設定された後、Altibaseはユーザのオペレーションを介入することなく自動的にレプリケーションを実行し、データの同一性が保証します。

Altibaseは、DB回復処理の高速化を実現するため、高速で安全なロギング処理も実行しています。

Concurrent Accessibility to Memory Table and Disk Table

Altibaseの一つの特徴として、前述した通りメモリテーブルとディスクテーブルを同時にデータ処理可能な点が挙げられます。ここでは、この機能について詳細に例を交えて説明します。図 4 のデータ定義で Account テーブルとTransaction 1 テーブルをメモリテーブル上に作成します。Transaction 2 テーブルと Customer テーブルはディスクテーブル上に作成します。Transaction 1 と 2 のテーブルは同じテーブル構造とします。これらの違いは直近のデータはメモリテーブルに(Transaction 1)に、それ以外のデータはディスクテーブル(Transaction 2)に 格納するという点です。この後、DML (Insert/Delete/Update) オペレーション、Joinオペレーション、サブクエリーオペレーションについて順に説明します。



<図 4> Account/Transaction DB Schema

DML (Data Manipulation Language)

メモリテーブル上の一か月以上前のデータ(Transaction 1)をディスクテーブル(Transaction 2)に移動します。テーブル間のデータの移動にはMOVE DMLを使用します。以下にMOVE DMLの実行例を示します。

```
MOVE INTO Transaction2 Transaction1 where date > 2004-10-16;
```

Altibase 4 supports flexible processing functionality between memory tables and disk tables.

Join

メモリテーブルとディスクテーブルを結合したい場合、以下のようにクエリーを記述します。(クエリーを記述する際には、メモリテーブルかディスクテーブルかを意識する必要はありません)

```
SELECT no FROM Customer, Account where Customer.ID = Account.ID;
```

Sub-query

ある顧客のトランザクションデータをサブクエリーで参照します。以下のように記述します。

```
SELECT * FROM Transaction2 D1 WHERE D1.no IN (SELECT Account.no FROM Account.ID = '731220-1148321');
```

Retrieval by View

同一スキーマのテーブルがメモリテーブルとディスクテーブル上に別々に定義されている場合、ビューを用いて一つのテーブルであるかのように処理す

することもできます。

```
CREATE VIEW Transaction * AS SELECT * FROM Transaction1  
UNION ALL SELECT * FROM Transaction2;
```

```
SELECT ID FROM Transaction WHERE date < 2004-11-17 AND  
date > 2004-10-10;
```

Altibaseはビュー処理の効率化にも重点を置いています。上記の例では、ビューは通常WHERE句で指定された条件をUNIONに反映させませんが、WHERE条件を二つのテーブル(Transaction1と2)に各々反映させ、index参照をさせることにより高速参照処理をするような工夫をしています。

Three Operating Models

Altibaseは様々なDB環境に適応するために次の3つのオペレーティングモデルを提供しています。

MMDB Model

高速処理が求められ、DB全体をメモリに格納できる十分なサイズの物理メモリを調達できる場合、全てのテーブルをAltibaseのメモリテーブルスペース上に定義します。このモデルではディスクテーブルを定義しません。このDBモデルの良い事例は通信分野のシステム、ソフトスイッチ、HLR、DLRなどです。

DRDB Model

DB処理性能よりも大容量データを格納するDBとして利用する場合のモデルです。このモデルではメモリテーブルを定義せずにディスクテーブルのみ定義します。社内システムのDBとして運用する場合などに利用します。

Hybrid DB Model

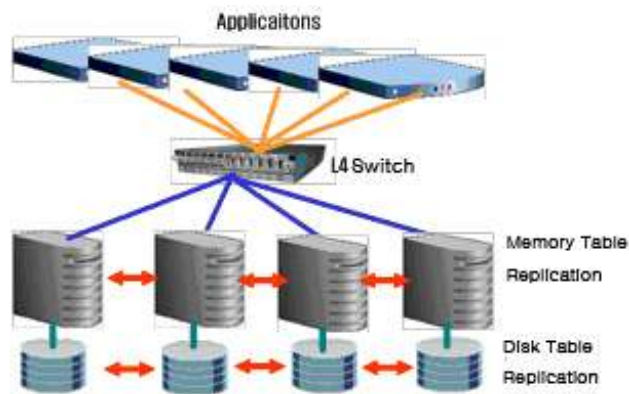
大容量のデータ格納要件もあり、さらに高速にデータ処理する必要がある場合にこのモデルが利用されます。ハイブリッドモデルではホットデータ (頻繁にアクセスされるデータ) をメモリテーブルスペース上に定義し、コールドデータ(アクセス頻度の低いデータ)はディスクテーブルスペース上に定義します。

Complete DB Recovery

Altibaseは高速なチェックポイント取得が可能です。DB障害が発生した場合、過去のチェックポイント処理状況を把握し、最後に取得したチェックポイントファイルから回復処理を実行します。従って、システム回復時間も最小の時間で済みます。DB回復処理ではローカルログの完全性もチェックされます。

DB Replication

サービス中にDBに障害が発生した場合、Altibaseはリモートのサーバに同期的にデータをレプリケーションしているので迅速なDB切り替えが可能です。Altibaseのレプリケーションコストは非常に小さく、スタンドアロン時のトランザクション処理性能をほぼキープします。



<図 5> Altibase Replication

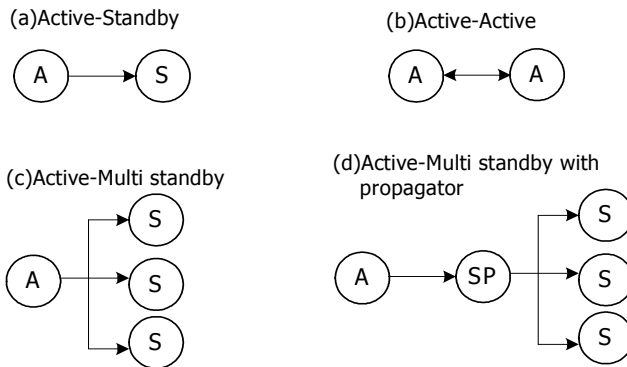
Altibase のレプリケーションは、テーブル単位で定義可能です。メモリテーブルとディスクテーブル間のレプリケーションも可能です。図 6でレプリケーション機能を説明します。

ログによるレプリケーション方式

ローカルサーバの性能がリモートサーバに影響を与えない方式

1 to N のレプリケーションをサポート

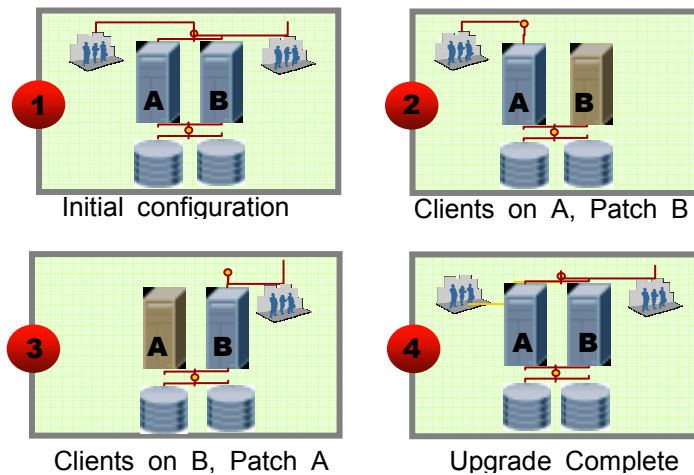
図6に示すような様々なレプリケーションを構成可能



<図 6> Replication Architecture

Upgrade without Shutdown

24時間サービスへ対応するため、DBシステムは障害回復の仕組みを持っている必要があります。DBMSサービス中にソフトウェアを入れ替えるのは単純ではなく、管理者の手を煩わせます。Altibaseではレプリケーション構成をとっている場合、システム全体をシャットダウンすることなく(サービスを停止することなく)アップグレード処理が実行可能です。



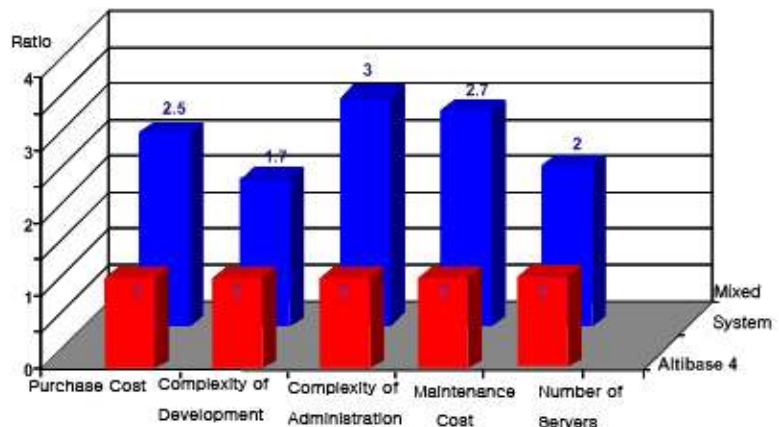
<図 7> Procedure or upgrade without shutdown

High Performance with Low Investment Cost

Investment Cost

大容量で高速でかつ信頼性の高いDBシステムを構築する場合、Altibaseを

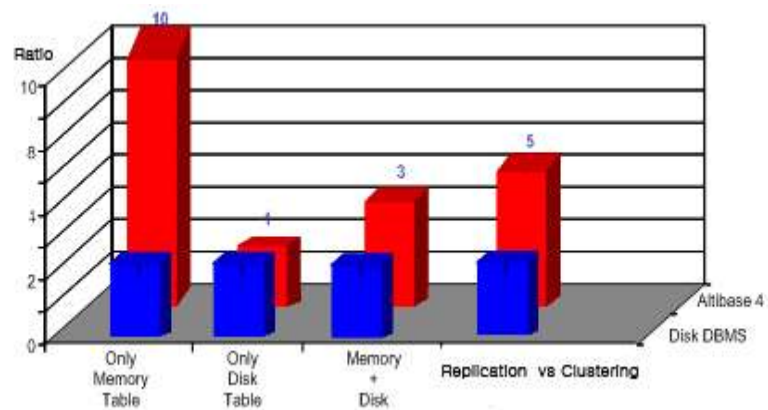
単体で使用する構成とMMDBMSとDRDBMSを両方組み合わせる構成をいくつかの観点で比較しておくべきです。まずコスト面においては、S/W費用、H/W費用、開発費用、運用・保守費用などです。S/W費用はAltibase単体の場合とMMDBMSとDRDBMSを組み合わせた場合と比較すると二つの製品を購入する必要があるため後者が高くなります。開発費用はアプリケーションが二つのDBを意識する必要がある点、二つのDBMS間でのデータ同期などが必要なことから後者が高くなります。さらに運用・保守費用についても二つのDBを運用管理するコストがあり後者が高くなります。同一性能を引き出すという前提で考えた場合、Altibaseを単体で使用した場合と比較してMMDBMSとDRDBMSを組み合わせた構成の場合は図8に示す通り2倍以上のコストが必要です。



<図 8> Comparison of Cost between Altibase and Mixed System

Performance

ディスクDBとメモリDBをサポートするAltibaseはDRDBMSよりも性能面で優位です。図 9は、AltibaseとDRDBMSとの性能を比較したものです。さらにAltibaseレプリケーション構成とDRDBMSのクラスタ構成での性能比較も示しています。AltibaseのメモリDBのみのモデルでは、DRDBMSの約10倍であり、AltibaseのディスクDBはディスクDBモデルと同等の性能を示します。Altibaseのハイブリッドモードでも約3倍の性能です。AltibaseのレプリケーションにおいてもディスクDBのクラスタ構成と比較して約5倍の性能です。



<図 9> Comparison of Performance between Altibase and Disk-Based DBMS

Conclusion

DBシステムを構築する際、様々な選択肢の中から有効で最適な構成を選択決定することはポイントの一つです。しかし、全てのあらゆるニーズに対応できるDBMSは市場にはなく、それを探するのは現時点において非常に困難だといえます。Altibaseはこれまで述べてきた通り新しいコンセプトのDBMSであり、以下の領域で様々なニーズに応えます。

- 1) 高性能、高可用性、メモリDBでは扱えない大容量データ管理が求められる分野
- 2) 小規模DBで最速の性能が求められる分野
- 3) 大規模DBであるが特に性能優先度は高くない分野

Altibaseは、メモリDBとディスクDBを組み合わせ一つにしたハイブリッドDBMSです。Altibaseを選択することで、低コストで効率的なDBシステムを構築することが可能になります。