

InterSystems ホワイト・ペーパー

パフォーマンス・ケース・スタディ

Mary Finn
Product Marketing Manager
InterSystems Corporation

概要

データベース製品比較の最良の方法は、実際のアプリケーション、できれば今稼働しているもので、直接的にテストすることです。特にポストリレーショナル技術の Caché を評価する場合には、これが最良の方法です。なぜなら、「標準的な」トランザクション処理のベンチマーク手法は、リレーショナルデータベースの「行と列」形式の制限を受けているからです。それらは、複雑なデータ・モデルをもつ実際のアプリケーションの正確な性能を予測することはできません。

よく知られたデータベース・ベンダが契約上で禁止しているため、Oracle や Microsoft SQLサーバなどのリレーショナル・データベースを使った実世界の性能テストは、ほとんど公表されていません。自社でそういった検証を行う企業がベンチマーク・テストを行うと、多くの場合、アプリケーション・コードの変更をせず、SQLのみを使用した場合でも、Caché が、リレーショナル・データベースよりも5倍程度高速であると分かります。この性能の違いは、さらに、Caché のポストリレーショナル技術の優位性を最大化させたアプリケーションでは、さらに増大します。

はじめに

多くの企業は Caché と他の多くのデータベースの直接的な比較を行っています。この文書は、独立系アナリストやInterSystems の顧客によるパフォーマンス及びスケーラビリティのベンチマークと、評価されたそれぞれのデータベース技術のオプションを基にして調査を行った結果です。上述した契約上の制限により、パフォーマンスのベンチマークを実施した企業名及び、Caché の競合するリレーショナル・データベースを取り扱う企業名は、ここでは省略します。

このテストは実世界の条件下で、実際のアプリケーション、および実際のデータを使用して実施されており、結果は最終的なものです。データベースへの問い合わせに SQL を使用する場合、Caché は通常、Oracle や SQL Server など他のリレーショナル・データベースよりも5倍あるいはそれ以上優れた性能を発揮します。また Caché は、Caché の多次元データ構造に直接アクセスするためにコードが最適化されるときは、さらに高速に実行します。

ここに上げる多くのケース・スタディでは、クエリ言語としてSQLを使用しています。理由は、より簡単に Caché とリレーショナル・データベースを直接的にテストできるためです。これらのケースでは、Caché の多次元構造により、複数のテーブルのジョインにかかる処理のオーバーヘッドがないため、優れた性能を示しています。多次元構造は、複数のテーブルのジョインにかかる処理のオーバーヘッドがありません。Caché は、Caché多次元データ構造に直接アクセスするようコードを最適化すれば、さらに数倍も高速になる可能性があります。

ケース・スタディ #1

ある電力供給会社が、その会社で使用しているデータウェアハウジング ETL (Extract - 抽出、Transform - 変換、Load - ロード) で、有名なリレーショナル・データベースよりも Caché の方が 5 倍も高速であることがわかりました。電力会社の社員は、1300 万件以上のソース・レコードを処理しながら、アプリケーションのフィールド・メンテナンス・モジュールを使用して、両方のデータベースをテストしました。両方のデータベースとも、同じテーブル定義とクエリを使用します。しかし既存のフィールド・メンテナンス・モジュールは、データベース特有のストアド・プロシージャに高度に依存しているので、このプロシージャは Caché のストアド・プロシージャ言語を使用して再実装されました。(これには、40時人のみ必要でした)

以下の表は、ベンチマークの結果をまとめたものです。

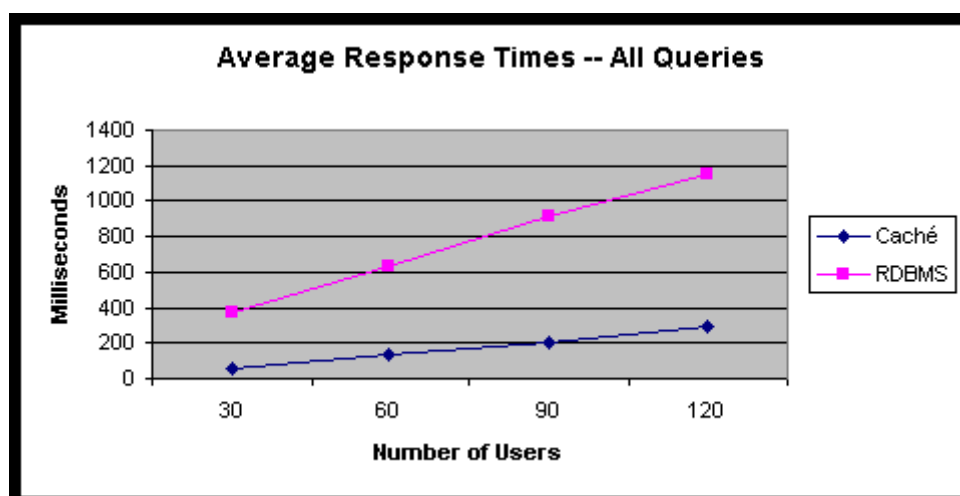
タスク	経過時間		パフォーマンス対比 Caché : RDBMS
	RDBMS	Caché	
ステージテーブルヘデータをロードする	138 分	6 分	23 : 1
2 番目のステージテーブルにデータをロードし、データをフィルタする	245 分	23 分	10 : 1
フィールド・メンテナンスの次元テーブルを生成する	168 分	24 分	7 : 1
フィールド・メンテナンスのファクト・テーブルを生成する	890 分	233 分	4 : 1
データのロード、およびデータのフィルタの合計時間	1441 分	286 分	5 : 1

ケース・スタディ #2

あるヘルスケア・アプリケーション・プロバイダは、Caché と有名なリレーショナル・データベースの SQL 応答時間を比較するテストを実施しました。これは、これまでの患者データ (7 つのテーブル、および 650 万件以上のレコード) を使用して、8 つのクエリー式は 30、60、90、および 120 の同時ユーザの擬似的にロードして実行されたものです。

以下のテーブルは、両方のデータベースに対する平均応答時間をミリ秒単位で表しています。両方のデータベースとも高速ですが、Caché は平均 4 倍から 6 倍も高速です。また、結果のグラフは Caché がよりスケーラビリティに優れていることを示しています。例えば、ロードが 90 から 120 ユーザに拡張されるとき、Caché の応答時間は 84 ミリ秒増えるだけで済むのに対し、他のリレーショナル・データベースは 231 ミリ秒 (Caché の 2.7 倍) 遅くなります。

同時 ユーザ数	平均応答時間 (8 つのクエリ全体にわたって)		パフォーマンス対比 Caché : RDBMS
	RDBMS	Caché	
30	375.125 ミリ秒	59.125 ミリ秒	6.3 : 1
60	637.25 ミリ秒	137.75 ミリ秒	4.6 : 1
90	915.625 ミリ秒	206.875 ミリ秒	4.2 : 1
120	1146.375 ミリ秒	290.125 ミリ秒	3.9 : 1



ケース・スタディ #3

あるコンサルタントは、データ・ウェアハウジング・シナリオにおけるパフォーマンスを計測するため、有名なリレーショナル・データベースと Caché を比較しました。これは、100 万件のレコード・テーブルの実際のオーダー処理データを使用して、ロードのパフォーマンスとクエリの応答時間を計測したものです。

Caché のデータ構造はパフォーマンスに対して最適化されていなかったものの、1,000,000 件のロードは、Caché は RDBMS の 3 倍(2681 秒 vs. 8567 秒)の速度で処理が完了しました。また、計測後のディスク容量は RDBMS は 216 MB 必要とするのに対し、Caché データベースは146 MB しか必要としませんでした。

以下の表は、SQL クエリの応答時間の結果です。

クエリ	応答時間		パフォーマンス対比 Caché : RDBMS
	RDBMS	Caché	
全ての次元に対するテーブル 全体の大量の集計	638 秒	180 秒	3.5 : 1
全ての次元の集計(2つの次元 については制限付き)	3 秒	1 秒以下	>3 : 1
全テーブルにおける未使用の Purpose コードの検索	8 秒	3 秒	4 : 1
ある次元が指定された数よりも大 きい場合の全レコード検索	50 秒	16 秒	3.1 : 1

ケース・スタディ #4

全国規模の電話会社は、独自のオンライン電話帳アプリケーションを使用したパフォーマンス・テストで、Cache が有名なリレーショナル・データベースよりも平均 10 倍の速度で処理できるということがわかりました。

テキスト・ファイルからリレーショナル・データベースをロードすると約 10 時間かかり、ディスク容量は 60 GB 必要としていましたが、Cache を使用した場合は約 45 分で済み、ディスク容量は 2 GB のみで十分です。Cache データベースは容量の使用において劇的に効率的になったので、キャッシュ・メモリの使用はより改善されました。その結果、クエリの応答時間は RDBMS よりも Cache の方がより高速になったのです。

クエリ	応答時間		パフォーマンス対比 Cache : RDBMS
	RDBMS	Cache	
名前、住所による検索	0.33 秒	0.07 秒	4 : 1
名前による検索	0.46 秒	0.02 秒	23 : 1
名前、住所、郵便番号による検索	0.76 秒	0.04 秒	19 : 1
番地による検索	0.14 秒	0.02 秒	7 : 1
名前1、名前2、住所、および郵便番号による検索	0.01 秒	0.01 秒	1 : 1
国名による検索	0.02 秒	0.01 秒	2 : 1
全クエリの平均時間	0.287 秒	0.028 秒	10 : 1

インデックスは、クエリのパフォーマンスを向上させるために RDBMS でも Cache でも使用されてきましたが、RDBMS は Cache よりも低速でしたが、それでも 1 秒以下の応答時間を維持しています。しかし、RDBMS への更新（および必要なインデックスの再構築）は、重大なパフォーマンスの問題を避けるためにオフライン時に実施しなければなりません。実際にはアプリケーションは、2 つのリレーショナル・データベースを使用するように設計されています。1 つがライブ状態であるとき、もう 1 つは更新中で、この状態は 1 日ずつ交替します。対照的に、ライブ状態である Cache データベースは、パフォーマンスの低下などの重大な問題が発生することなく更新することができます。また Cache では、RDBMS にはない 前方一致検索が使用できます。

ケース・スタディ#5

通信分野のソリューションプロバイダは、ショートメッセージ・システム(SMS)のイベント追跡アプリケーションの開発を依頼されました。アプリケーションは、複数のリモートソースからの大量データを受け、レコードのインデックスを付して、ディスク上の中央データベースに書き込むことを求められました。アプリケーションのプロトタイプを使い、そのソリューションプロバイダは、Cache とよく知られたリレーショナル・データベースとの比較を行いました。

初期テストにおいては、両方のデータベースで、同じコードを使用しました。システムには他の処理ロードはありません。その結果は以下に示す通りです。

	平均 レコード/プロセス秒		パフォーマンス対比 Cache : RDBMS
	RDBMS	Cache	
ロード、インデックス、データ書き込み	315	1700	5 : 1

この結果に基づき、RDBMSによるアプリケーションの開発は中止されました。コードをCacheのネイティブ多次元データ構造向けに最適化すると、アプリケーションの性能が6倍に上がりました。マルチプルストリームインプットを使用して、システムの処理ロードをシミュレーションしたところ、最適化したCacheベースのアプリケーションは、平均で1秒間に10,600レコードの処理スピードを達成しました。

まとめ

データベース製品比較の最良の方法は、実際のアプリケーション、できれば今稼働しているもので、直接的にテストすることです。このホワイト・ペーパーに記したケース・スタディでは、Cachéとよく知られたリレーショナル・データベースとの直接的な性能テストを示しています。このテストは、独立アナリストとインターシステムズ社のお客様やご検討いただいているお客様が、実際のアプリケーションで実施したものです。その結果、Caché は、リレーショナル・データベースよりも 5倍高速にSQLクエリを処理することが分かります。また、Cachéの多次元データ構造を直接操作するようコードを書き換えれば、さらに高速性を実現することが可能となります。

インターシステムズジャパン株式会社

〒160-0023

東京都新宿区西新宿6-10-1 日土地西新宿ビル 17F

TEL: 03-5321-6200 (代)

InterSystems.co.jp

INTERSYSTEMS

InterSystems Ensemble® と InterSystems Caché®は、米国インターシステムズ社および、その子会社の登録商標です。米国インターシステムズ社およびその子会社の商標です。その他の製品名は、
該各社の商標または登録商標です。Copyright © 2011 InterSystems Corporation. All rights reserved. 11-08.