

Innovations by InterSystems

InterSystems のテクノロジーがもたらす
ビジネス・インテリジェンス・ソリューション

技術白書:

InterSystems Corporation,
Senior Sales Engineer, Dr. Mark Massias

エグゼクティブ・サマリ

ビジネス・インテリジェンス・ソリューションのトレンドは、データ遅延を減らし、クエリに対してよりタイムリーに応答することです。このトレンドが積み重なって、まったく新しい種類のアプリケーションが生まれました。"ビジネス・インテリジェント・アプリケーション" では、リアル・タイムで運用データを分析および処理できます。ビジネス・アクティビティ・モニタリング (BAM) ソリューションは、統合システムに対して同様のことを行います。

InterSystems Caché データベースには、リレーショナル・データベースと比較した場合、5 倍の速度の SQL パフォーマンスを持つ多次元データ・アーキテクチャが使用されています。従来のデータ・ウェアハウス・ソリューション、および "トランザクション・ビットマップ・インデックス" やリアルタイム・データ解析の実現に、Caché のハイパフォーマンスは最適です。

InterSystems の共通統合プラットフォームである Ensemble は、同様のリアルタイム・データ分析機能を統合システムに提供し、ビジネス・アクティビティ・モニタリング・ソリューションの構築を容易にします。

はじめに - ビジネス・インテリジェンスの現況

さまざまな IT システムで日々処理されるデータは、タイムリーに保存および分析され、役に立つような情報に変換された場合、多くの企業にとって莫大な戦術的および戦略的価値があるものとなります。これが "ビジネス・インテリジェンス" アプリケーションが果たす役割であり、その構築、メンテナンス、向上に多くの組織は莫大な労力を注いでいます。

ビジネス・インテリジェンス・アプリケーションの最も一般的なアーキテクチャには、データ・ウェアハウスまたはデータ・マートに使用されるリレーショナル・データベースが含まれます。これには通常、1 つ以上の "トランザクション" システムからのデータが使用され、バッチ・モードで定期的にアップデートされます。蓄積されたデータは、"データ・マイニング" およびアドホック・クエリをサポートする多数のレポート・ツールによってアクセスおよび分析されます。

しかし、多くの企業は、リレーショナル・データ・ウェアハウスにはいくつかの欠点があることに気付いています。複雑なデータをうまく操作できないというのがその 1 つです。また、データ遅延も欠点の 1 つです。アップデートの頻度によって、ウェアハウス内のデータは数時間、または数日遅いものである場合があります。

したがって、ビジネス・インテリジェンス・テクノロジーのトレンドとして、より迅速で、より頻繁にアップデートされるデータ・ウェアハウスまたはデータ・マートを構築することに力が注がれてきました。これらのトレンドが積み重なると、当然、集大成はスタンドアロンのトランザクション・システムからリアルタイムでデータを分析および処理する性能を持つ新種のアプリケーションとなります。

リアルタイム・データ分析は統合システムにも採用できます。ビジネス・アクティビティ・モニタリング (BAM) ソリューションの形式には、通常、組織中から集められたさまざまなメトリックの現在値を表示する "エグゼクティブ・ダッシュボード" が使用されます。

この白書では、多次元データベースである InterSystems Caché を使用して、ハイパフォーマンス・データ・ウェアハウスおよび "ビジネス・インテリジェント・アプリケーション" を構築する方法の概要を説明します。このスタンドアロン・システムでは、情報がリアルタイムで提供され、景況の変化に瞬時に対応します。また、統合システムでビジネス・アクティビティ・モニタリングを実行するための選択肢として、InterSystems の統合プラットフォームを最適なものにする Ensemble の機能について説明します。

ハイパフォーマンス・データ・ウェアハウスとしての Caché

Caché は、リレーショナル・データベースの主要な利点を持ちますが、リレーショナル・データベースではありません。Caché のデータ構造の基本はスパース多次元配列です。これはリレーショナル・データベースで使用されるテーブル構造より、特にそれが複雑である場合、データの保存に効率的です。Caché はリレーショナル・データベースの特徴である "ジョイン" や "テーブルホッピング" の制約を受けないので、処理のオーバーヘッドが少なく、大幅に早くクエリに応答できます。

Caché はリレーショナル・データベースではありませんが、ODBC および JDBC 対応のクエリ言語を使用できます。これは Caché の多次元データが、リレーショナル・テーブルとして自動的に表示されるからです。したがって、Caché はリレーショナル・データベース用に設計されている分析およびレポート・ツールの多くにアクセスできます。実際のアプリケーションを使用しての直接テスト¹ では、Caché はリレーショナル・データベースより、通常 5 倍早く SQL クエリに応答することが示されました。また、アップデート時にウェアハウスにデータを挿入するのもより迅速です。

¹詳細は、InterSystems 技術白書、「パフォーマンス・ケーススタディ」を参照してください。

Caché データベースの多次元データ構造に直接アクセスするときも、よりよいパフォーマンスを得ることができます。

市販のいくつかのデータ分析およびレポート・ツールは、Caché に直接多次元アクセスします。

Caché ベースのデータ・ウェアハウスは高速であるだけでなく、同じデータを保持するリレーショナル・ウェアハウスよりもリソースの消費が少ないです。多次元行列のスパースのために、Caché が使用するハード・ディスク容量は、リレーショナル・システムと比較し、1/2 から 2/3 少なくなっています。また、チューニングが不要なので、管理のオーバーヘッドも少なくて済みます。

データ遅延の削減

ビジネス・インテリジェンス・アプリケーションの中には、特定の時間のデータ遅延が許容されるものもあります(月次の売上傾向を調査する経営分析者は、データがそれほど頻繁にアップデートされなくても満足するでしょう。また、複雑なクエリの処理に数秒かかったとしても問題視しません)。しかし他のアプリケーション(不正防止またはクロス・セリングを目的とする、現在と過去の購買傾向を比較するアプリケーションなど)では、最新のデータで処理を行う必要があります。

データ遅延を削減する一般的な方法の 1 つに、シャド・サーバ・テクノロジーの使用があります。基本(運用)データ・ストアのデータに変更があった場合、それらのトランザクションはジャーナル・ファイルに書き込まれます。ジャーナル・ファイルは 1 台上のシャド・サーバに読み取られ、変更はデータベースの複製に適用されます。クエリはシャド・データベースに対して実行され、基本マシンのトランザクション・パフォーマンスを落としません。データ遅延は、ジャーナルに記録されたトランザクションを読み取り適用する時間に削減されます。シャド・サーバの処理中のクエリの数に応じ、データ遅延は数秒ほどになります。

Caché ではシャド・サーバ・テクノロジーがサポートされています。ハイパフォーマンスであるために、シャド・データベースのアップデートおよびクエリの応答にかかる時間が削減されます。

Caché でのリアルタイム・データ分析

場合によっては、シャド・システム側での数秒のデータ遅延も遅い場合があります。場合によっては、急速に変化する運用データに直接クエリを実行する必要があります。このような場合には、アプリケーション全体のパフォーマンスを落とさないために、極めて高速にクエリが処理される必要があります。

いかなるデータ分析システムにおいても、データベース検索の基準として頻繁に使用されるプロパティにインデックスを作成することが、クエリの応答時間を向上させる 1 つの方法です。Caché では、開発者が従来のインデックスおよびビットマップ・インデックスを構築できます。

従来のインデックスでは、プロパティ（リレーショナル用語ではテーブル内の列）はレコード ID の一覧です。各プロパティ値に対してレコード ID 一覧があり、インデックス付きのプロパティがその値とマッチするようになっています。

The diagram shows a table titled "Traditional Index on Property: 'Type of Pet'". It lists two categories: "Cat" and "Iguana". For "Cat", the associated record IDs are 5, 8, 24, 25, 34, 55, 77, 102, and 104, followed by an ellipsis. For "Iguana", the record IDs are 2, 333, 1568, 6703, 8932, and 10982, followed by an ellipsis. A vertical ellipsis is shown below the "Iguana" row.

Property Value	Record IDs
Cat	5, 8, 24, 25, 34, 55, 77, 102, 104 ...
Iguana	2, 333, 1568, 6703, 8932, 10982 ...

ビットマップ・インデックスは、大量のデータの分析速度を大幅に向上させるので、データウェアハウス・ソリューションでよく使用されます。ビットマップ・インデックスでは、プロパティはビット文字列です。どのプロパティ値にも、各レコードを示すビット文字列があります。レコードにそのプロパティ値がある場合、ビットは "1" で、ない場合は "0" です。

The diagram shows a table titled "Bit-Map Index on Property: 'Hair Color'". It lists two categories: "Black" and "Red". For "Black", the bit string is 001011000010110010100101100, followed by an ellipsis. For "Red", the bit string is 010000011001001000010000010, followed by an ellipsis. A vertical ellipsis is shown below the "Red" row.

Property Value	Bit String
Black	001011000010110010100101100 ...
Red	010000011001001000010000010 ...

ビットマップ・インデックスの利点は、インデックスに対してブーリアン処理 (AND、OR) を実行してクエリを処理できることです。これにより、データベース全体を検索することなく、どのレコードがクエリ条件に正確に一致するか効率的に判断できます。複雑なクエリの場合、ビットマップ・インデックスでは 100 倍の率で応答時間を削減できます。

ただし、既存のデータが頻繁に変更または削除される場合、ビットマップ・インデックスは簡単にアップデートできません。このために、ビットマップ・インデックスは、今まで静的な読み取りのみのデータ（データウェアハウス・ソリューションなど）が使用される状況においてのみ使用

されてきました。またビットマップ・インデックスは大量のハードディスク容量を必要とします。インデックス内のすべての値には各レコードに対するビット (1 または 0) があるからです。

Caché は "トランザクション・ビットマップ・インデックス" を実現する唯一のデータベースです。その効率的な多次元データ構造に高度な圧縮技術が組み合うことにより、Caché のビットマップ・インデックスは、他のデータベース・ベンダが提供するよりも小さく、また高速にアップデートできます。Caché のビットマップ・インデックスは従来のものとアップデートに要する時間が変わらないので、迅速に変化するデータで使用するのに適しています。

開発者は、トランザクション・ビットマップ・インデックスにより、迅速に変化する運用データを分析し、それをリアルタイムで処理することができる "ビジネス・インテリジェント・アプリケーション" を構築できます。

統合システムにおける Ensemble でのビジネス・アクティビティ・モニタリング

ビジネス・アクティビティ・モニタリング (BAM) 機能を持つ統合システムにおいても、リアルタイムでのデータ処理が求められています。BAM ソリューションでは、瞬時に収集および分析される情報はさまざまな異なるソースから来ます。

InterSystems の共通統合プラットフォームである Ensemble には、組み込み型のデータ・リポジトリがあり、統合システムを通過するすべてのメッセージはすべてそこに保存されます。この "メッセージ・ウェアハウス" は Caché と同じ多次元のデータ構造を使用し、同じ極めて高速なパフォーマンスを提供します。Ensemble はリアルタイムでウェアハウス内のデータを分析でき、強力なデバッグ、管理および BAM 機能を実現します。

Ensemble には BAM ソリューションの開発を高速化するように設計されている機能があります。Ensemble の高度なオブジェクト・テクノロジーとウィザード形式のスタジオは、開発者が定義した間隔で情報を収集あるいは計算する "ビジネス・メトリック" の定義を容易にします。その情報は、カスタマイズ可能な "エグゼクティブ・ダッシュボード" に表示される計器と関連付けることができます。また、統合システム内にフィードバック・ループを実装するロジックを、ビジネス・メトリックに含めることができます。例えば、測定パラメータが予測を超えた場合、ビジネス・メトリックによって自動的に修正や警告が行われます。

まとめ

InterSystems の製品である Caché (ハイパフォーマンス・データベース) および Ensemble (共通統合プラットフォーム) は、ビジネス・インテリジェンスのソリューションに適した多次元データ構造を使用します。従来のデータ・ウェアハウス・アプリケーションは、極めて高速な Caché のパフォーマンスの利益を享受できます。しかしさらに重要なことに、このハイパフォーマンスによって、迅速に変化する運用データをリアルタイムに分析することが可能になり、それが新種のビジネス・インテリジェンス・ソリューションの可能性を広げます。"ビジネス・インテリジェント・アプリケーション" は、リアルタイムで情報を分析および応答するスタンドアロン・システムであり、これにより企業は、日々の業務の一環として収集されるデータを完全に活用できるようになります。

Ensemble は、統合システム内のメッセージやその他のデータのリアルタイム分析を実現します。Ensemble では、ビジネス・アクティビティ・モニタリング・ソリューションで必要とされる、エグゼクティブ・ダッシュボードやフィードバック・メカニズムを簡単に構築できます。

* 本白書は、米国インターシステムズ社の "How InterSystems Technology Enables Business Intelligence Solutions" の日本語訳です。不明な点は、英文本文にてご確認ください。

INTERSYSTEMS

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 日土地新宿ビル17階

www.intersystems.co.jp

InterSystems Caché、InterSystems Ensemble は、米国インターシステムズ社の商標です。

他の製品名は各社の商標です。

© copyright 2005 InterSystems Corporation. All rights reserved 10-05