

INTERSYSTEMS **WHITE PAPER**

BUSINESS INTELLIGENCE-LÖSUNGEN MIT INTERSYSTEMS-TECHNOLOGIE

Ein Whitepaper von:

Dr. Mark Massias

Senior Sales Engineer

InterSystems Corporation

The logo for InterSystems, featuring the word "INTERSYSTEMS" in a stylized, blue, serif font. The letters are closely spaced and have a slightly irregular, hand-drawn appearance.

Management Summary

Die Entwicklung bei Business Intelligence-Lösungen geht dahin, diese zeitnaher zu gestalten und auf diese Weise sowohl die Datenlatenz wie auch die Reaktionszeit bei Abfragen senken. Den Höhepunkt dieser Entwicklung bildet eine ganz neue Art von Anwendungen: "Intelligente Geschäftsanwendungen" können operative Daten in Echtzeit analysieren. Business Activity Monitoring-Lösungen übernehmen dasselbe für integrierte Systeme.

Die Datenbank Caché von InterSystems basiert auf einer multidimensionalen Datenarchitektur, die eine bis zu fünfmal schnellere SQL-Performance als relationale Datenbanken bietet. Caché ist auf Grund seiner hohen Performance die ideale Lösung für herkömmliche Data Warehouse-Lösungen und zeichnet sich aus durch "transaktionale Bitmap-Indizierung" und Datenanalysen in Echtzeit.

Die universelle Integrationsplattform Ensemble von InterSystems bietet dieselben Technologien zur Datenanalyse in Echtzeit und vereinfacht so die Erstellung von Business Activity Monitoring-Lösungen.

Business Intelligence im aktuellen Einsatz – Eine Einführung

Für die meisten Unternehmen haben die tagtäglich von den verschiedenen IT-Systemen verarbeiteten Daten enormen operativen und strategischen Wert – vorausgesetzt, sie können gespeichert, analysiert und zeitnah in nützliche Informationen umgewandelt werden. Dies ist die Aufgabe von "Business Intelligence"-Anwendungen. Viele Organisationen investieren viel Arbeit in deren Erstellung, Pflege und Verbesserung.

Die gängigste Architektur einer Business Intelligence-Anwendung umfasst eine relationale Datenbank, die als Data Warehouse eingesetzt wird und in regelmäßigen Abständen mit Informationen aus einem oder mehreren "transaktionalen" Systemen aktualisiert wird. Auf die angesammelten Daten wird zur Analyse über verschiedene Auswertungs-Tools zugegriffen, wobei vielfach "Data-Mining" und Ad-hoc-Abfragen durchgeführt werden. In vielen Unternehmen wächst die Erkenntnis, dass relationale Data Warehouses auch einige Nachteile aufweisen: beispielsweise ungenügende Performance bei der Bearbeitung komplexer Daten oder die Latenz von Informationen. Je nachdem, wie häufig ein Warehouse aktualisiert wird, können die Daten Stunden oder sogar Tage alt sein.

Daher geht der Trend bei Business Intelligence-Technologie in Richtung Erstellung schnell verfügbarer, häufig aktualisierter Daten in Data Warehouses. Die logische Weiterentwicklung dieses Trends ist eine neue Art von Anwendung mit der Fähigkeit, Daten aus eigenständigen transaktionalen Systemen in Echtzeit zu analysieren und mit ihnen zu arbeiten. Die Datenanalyse in Echtzeit kann auch in integrierten Systemen zur Verfügung gestellt werden. Business Activity Monitoring (BAM)-Lösungen werden in der Regel als "Executive Dashboards" bereitgestellt, die aktuelle Werte der verschiedenen Metriken im Unternehmen darstellen.

Im vorliegenden White Paper wird beschrieben, wie mit der multidimensionalen Datenbank Caché von InterSystems hochperformante Data Warehouses und "Intelligente Geschäftsanwendungen" in eigenständigen Systemen erstellt werden können, die Informationen in Echtzeit zur Verfügung stellen und sofort auf sich ändernde Geschäftsbedingungen reagieren können. Darüber hinaus werden die Leistungsmerkmale erklärt, weswegen die universelle Integrationsplattform Ensemble von InterSystems, die bevorzugte Lösung für Business Activity Monitoring in integrierten Systemen ist.

Caché als hochperformantes Data Warehouse

Caché ist keine relationale Datenbank, obwohl sie alle wesentlichen Vorteile von relationalen Datenbanken bietet. Die zugrunde liegenden Datenstrukturen von Caché sind Sparse Arrays, die eine sehr viel effizientere Datenspeicherung besonders bei komplexen Daten erlauben, als die in relationalen Datenbanken verwendeten tabellenartigen Strukturen. Da bei Caché keine "Joins" und häufige Wechsel zwischen Tabellen wie bei relationalen Tabellen vorkommen, besteht weniger Verarbeitungs-Overhead und Abfragen können signifikant schneller beantwortet werden.

Obwohl Caché keine relationale Datenbank ist, unterstützt es als Abfragesprache ODBC- und JDBC-kompatible SQL. Denn die multidimensionalen Daten von Caché werden automatisch als relationale Tabellen dargestellt. Daher können die vielen Analyse- und Reporting-Tools, die für die Verwendung mit relationalen Datenbanken entwickelt wurden, auch auf Caché zugreifen. Praxisvergleiche mit echten Anwendungen¹ zeigten, dass Caché in der Regel auf SQL-Abfragen fünfmal schneller antwortet als relationale Datenbanken. Die Geschwindigkeitsvorteile bestehen auch bei Aktualisierungen durch Einfügen von Daten in das Warehouse.

Durch den direkten Zugriff auf multidimensionale Datenstrukturen in einer Caché-Datenbank kann eine weitere Performanceverbesserung erzielt werden. Verschiedene kommerziell verfügbare Datenanalyse- und Reporting-Tools verwenden den direkten multidimensionalen Zugriff auf Caché.

Ein Caché-basiertes Data Warehouse ist nicht nur schneller, sondern belegt auch weniger Ressourcen als ein relationales Warehouse mit denselben Daten. Dank der Dünnbesiedelung der multidimensionalen Arrays belegt Caché in der Regel nur die Hälfte bis zu zwei Drittel des Festplattenspeichers eines relationalen Systems. Caché erfordert kein Tuning, was wiederum den administrativen Overhead reduziert.

Geringere Datenlatenz

Einige Business Intelligence-Anwendungen können einen bestimmten Betrag an Datenlatenz tolerieren. (Ein Business Analyst, der die monatliche Umsatzentwicklung verfolgt, ist mit gelegentlich aktualisierten Daten zufrieden. Es spielt auch keine Rolle, ob für die Verarbeitung einer komplexen Abfrage mehrere Sekunden benötigt werden.) Aber andere Anwendungen hingegen (beispielsweise zum Vergleich aktueller und historischer Kaufmuster zur Betrugsverhinderung oder für Cross-Selling) benötigen für ihre einwandfreie Funktion immer ganz aktuelle Daten.

Eine gängige Methode, die Datenlatenz zu verringern, ist die Verwendung von Shadow Server-Technologie. Wenn sich Daten im primären (operativen) Datenspeicher verändern, werden diese Transaktionen in eine Journaldatei geschrieben. Ein oder mehrere Shadow Server lesen die Journaldatei und wenden die Änderungen auf eine duplizierte Datenbank an. Abfragen können auf die Shadow-Datenbank ausgeführt werden, ohne dass die transaktionale Performance des Primärsystems darunter leidet. Die Datenlatenz wird auf die Zeitspanne reduziert, die zum Lesen und Anwenden der protokollierten Transaktionen erforderlich ist. Je nachdem, wie viele Abfragen der Shadow Server zudem verarbeitet, kann die Datenlatenz auch einige Sekunden betragen.

Caché unterstützt Shadow Server-Technologie. Durch die hohe Performance wird sowohl die Zeitspanne für die Aktualisierung der Shadow-Datenbank wie auch die Beantwortung von Abfragen minimiert.

¹ Siehe auch das InterSystems-White Paper "Case Studies in Performance".

Caché für Datenanalysen in Echtzeit

In manchen Fällen sind jedoch schon die wenigen Sekunden Datenlatenz der Shadow-Systeme zu viel. Manchmal müssen Abfragen direkt auf schnell sich ändernde operative Daten ausgeführt werden. Dabei muss die Abfrageverarbeitung extrem schnell durchgeführt werden, um die Verringerung der Anwendungs-Performance insgesamt zu vermeiden.

In jedem Datenanalysesystem kann die Reaktionszeit für Abfragen optimiert werden, wenn Indizes für Eigenschaften erstellt werden, die häufig als Suchkriterium für die Datenbank verwendet werden. Caché gibt Entwicklern die Option an die Hand, sowohl traditionelle wie auch Bitmap-Indizes zu erstellen.

In einem traditionellen Index wird eine Eigenschaft (in relationalen Begriffen: eine Tabellenspalte) durch Listen von Datensatz-IDs beschrieben. Für jeden Eigenschaftswert liegt eine Liste von Datensatz-IDs vor, für die die indizierte Eigenschaft diesem Wert entspricht.



Bitmap-Indizes werden häufig in Data Warehouse-Lösungen verwendet, da sie die Analyse großer Datenmengen erheblich beschleunigen. Bei einem Bitmap-Index wird jede Eigenschaft als Folge von Bits beschrieben. Für jeden Eigenschaftswert gibt es eine Bitfolge, die jeden Datensatz darstellt. Das Bit hat den Wert "1", wenn der Datensatz den Eigenschaftswert besitzt, andernfalls ist er "0".



Der Vorteil von Bitmap-Indizes liegt darin, dass Abfragen mit Hilfe von booleschen Operatoren (UND, ODER) auf die Indexwerte verarbeitet werden können. Hierbei wird effizient ermittelt, welche Datensätze die Abfragebedingungen erfüllen, ohne dass dazu die gesamte Datenbank durchsucht werden muss. Für komplexe Abfragen können die Bitmap-Indizes die Antwortzeiten um einen Faktor von über 100 senken.

Die Aktualisierung von Bitmap-Indizes ist jedoch schwierig, wenn bestehende Daten häufig geändert oder gelöscht werden. Aus diesem Grund wurden sie herkömmlicherweise nur in Situationen mit statischen Daten verwendet, die wie in Data Warehouse-Lösungen nur gelesen werden. Bitmap-Indizes können auch große Mengen an Plattenspeicher belegen, da jeder mögliche Wert im Index ein Bit (entweder 1 oder 0) für jeden Datensatz enthält.

Caché ist die einzige Datenbank, die die "transaktionale Bitmap-Indizierung" unterstützt.

Effiziente multidimensionale Datenstrukturen und ausgefeilte Komprimierungstechniken machen die Bitmap-Indizes von Caché kleiner und viel schneller bei Aktualisierungen als die anderer Lösungen auf dem Markt. Mit Caché dauert die Aktualisierung von Bitmap-Indizes nicht länger als bei den traditionellen Alternativen, was sie für sich schnell ändernde Daten prädestiniert. Die transaktionale Bitmap-Indizierung ermöglicht Entwicklern "intelligente Geschäftsanwendungen" zu erstellen, die schnell sich ändernde operative Daten analysieren und die Ergebnisse in Echtzeit nutzen können.

Business Activity Monitoring in integrierten Systemen mit Ensemble

Um Daten in Echtzeit zu analysieren, ist es in integrierten Systemen ebenso erforderlich, Business Activity Monitoring (BAM)-Technologien bereitzustellen. Im Fall von BAM-Lösungen werden die Informationen ohne Verzögerung aus einer Vielzahl heterogener Quellen gesammelt und analysiert.

Die universelle Integrationsplattform Ensemble von InterSystems, verfügt über ein integriertes Daten-Repository, in dem alle Nachrichten gespeichert werden, die das integrierte System durchlaufen. Dieses "Message Warehouse" verwendet dieselben multidimensionalen Datenstrukturen wie Caché und bietet dieselbe hohe Performance. Ensemble kann Daten im Message Warehouse in Echtzeit analysieren, was leistungsstarke Debugging- und Management-Funktionalität inklusive BAM ermöglicht.

Ensemble enthält Technologie, die die Entwicklung von BAM-Lösungen beschleunigt. Die erweiterte Objekttechnologie von Ensemble und das assistentengesteuerte Studio vereinfachen die Definition von "Business-Metriken". Diese sammeln Informationen über ein integriertes System oder berechnen sie auf andere Weise in vom Entwickler festgelegten Abständen. Diese Informationen können mit Messinstrumenten verknüpft werden, die in einem anpassbaren Web-basierten "Executive Dashboard" angezeigt werden. Business-Metriken können auch Logik enthalten, die eine Feedback-Schleife in dem integrierten System implementiert. Wenn beispielsweise ein gemessener Parameter außerhalb eines erwarteten Limits liegt, kann die Business-Metrik automatisch eine Korrekturmaßnahme ergreifen, eine Benachrichtigung senden usw.

Zusammenfassung

Die hochperformante Datenbank Caché und die universelle Integrationsplattform Ensemble von InterSystems basieren auf einer mehrdimensionalen Datenarchitektur, die hervorragend für Business Intelligence-Lösungen geeignet ist. Herkömmliche Data Warehouse-Anwendungen können von der extrem schnellen Performance von Caché profitieren. Aber noch wichtiger ist, dass solch hohe Performance die Echtzeitanalyse von schnell sich ändernden operativen Daten ermöglicht. Dies ermöglicht neue Arten von Business Intelligence-Lösungen. "Intelligente Geschäftsanwendungen" sind eigenständige Systeme, die Informationen in Echtzeit analysieren und darauf reagieren können, so dass Unternehmen, die im Rahmen ihrer tagtäglichen Abläufe gesammelten Daten voll und ganz nutzen können.

Ensemble ermöglicht die Analyse von Nachrichten und anderen Daten in integrierten Systemen in Echtzeit. Mit Ensemble lassen sich die Executive Dashboards und Feedback-Mechanismen erstellen, die für Business Activity Monitoring-Lösungen erforderlich sind.

InterSystems GmbH
Hilpertstr. 20a
D-64295 Darmstadt
Tel.: +49.6151.1747-0
Fax: +49.6151.1747-11
www.InterSystems.de

INTERSYSTEMS