How InterSystems' Technology Enables Business Intelligence Solutions

A white paper by:

Dr. Mark Massias
Senior Sales Engineer
InterSystems Corporation

How InterSystems' Technology Enables Business Intelligence Solutions

Executive Summary

The trend in business intelligence solutions is to make them more timely – to decrease both data latency and the time it takes to respond to queries. The culmination of this trend is an entirely new kind of application. "Business intelligent applications" can analyze and act upon operational data in real time. Business Activity Monitoring solutions do the same for integrated systems.

InterSystems' Caché database uses a multidimensional data architecture that provides up to 5 times faster SQL performance than relational databases. Its high performance makes Caché ideal for traditional data warehouse solutions as well as enabling "transactional bit-map indexing" and real-time data analytics.

Ensemble, InterSystems' universal integration platform, provides the same real-time data analysis capabilities to integrated systems, making it possible to easily build Business Activity Monitoring solutions.

How InterSystems' Technology Enables Business Intelligence Solutions

Introduction – the State of Business Intelligence

For most enterprises, the data processed every day by their various IT systems has enormous tactical and strategic value – provided it can be saved, analyzed, and turned into useful information in a timely fashion. That is the role of "business intelligence" applications, and many organizations put a great deal of effort into building, maintaining, and improving them.

The most common architecture for a business intelligence application includes a relational database used as a data warehouse or data mart, which is periodically updated in batch mode with information from one or more "transactional" systems. The accumulated data can be accessed and analyzed by various reporting tools, many of which support "data-mining" and ad-hoc queries.

However, many organizations have discovered that relational data warehouses have some drawbacks. One is that their performance in manipulating complex data can be poor. Another is information latency. Depending on how frequently it is updated, the data in a warehouse may be hours, or even days, old.

Thus, the trend in business intelligence technology has been toward creating faster, more frequently updated data warehouses or data marts. The logical culmination of this trend will be a new kind of application with the capability to analyze and act upon data from standalone transactional systems in real time.

Real-time data analysis can also be employed in integrated systems. Business Activity Monitoring (BAM) solutions usually take the form of "executive dashboards" that display the current value of various metrics from across the enterprise.

This paper outlines how InterSystems' Caché, the multidimensional database, can be used to build high-performance data warehouses and "business intelligent applications", standalone systems that provide real-time information and respond instantly to changing business conditions. It will also discuss the capabilities that make Ensemble, InterSystems' universal integration platform, the ideal choice for business activity monitoring in integrated systems.

Caché as a High-performance Data Warehouse

Caché is not a relational database, although it offers all the key benefits of relational databases. Caché's fundamental data structures are sparse multidimensional arrays, which are much more efficient for storing data – particularly complex data – than the tabular structures used by relational databases. Because Caché is not subject to the "joins" and "table-hopping" that typify relational databases, it has less processing overhead and can respond to queries significantly faster.

Although Caché is not a relational database, it can use ODBC- and JDBC-compliant SQL as a query language. That's because Caché's multidimensional data is automatically exposed as relational tables. Therefore Caché is accessible to the many analysis and reporting tools that were designed for use with relational databases. Head-to-head tests using real applications* show that Caché typically responds to SQL queries five times faster than relational databases. It is also faster when inserting data into the warehouse during update procedures.

* See the InterSystems' white paper "Case Studies in Performance".

Even better performance can be achieved by directly accessing the multidimensional data structures in a Caché database. Several commercially available data analysis and reporting tools use direct multidimensional access to Caché.

In addition to being faster, a Caché-based data warehouse consumes fewer resources than a relational warehouse holding the same data. Thanks to the sparseness of its multidimensional arrays, Caché typically uses one-half to two-thirds less disk space than a relational system. And Caché doesn't require tuning, so its administrative overhead is less.

Reducing Data Latency

Some business intelligence applications can tolerate a certain amount of data latency. (The business analyst looking at month-to-month sales trends will be satisfied with data that is infrequently updated. Nor will it matter if it takes several seconds to process a complex query.) But other applications (for example, those that compare current and past buying patterns, either for fraud prevention or cross-selling opportunities) must be able to work with up-to-date data.

One common way of reducing data latency is to use shadow server technology. As data changes in the primary (operational) data store, those transactions are written to a journal file. One or more shadow servers read the journal file and apply the changes to a duplicate database. Queries can be run against the shadow database without degrading the transactional performance of the primary machine. Data latency is reduced to the amount of time it takes to read and apply the journalled transactions. Depending on how many queries the shadow server is also processing, the data latency may be as much as several seconds.
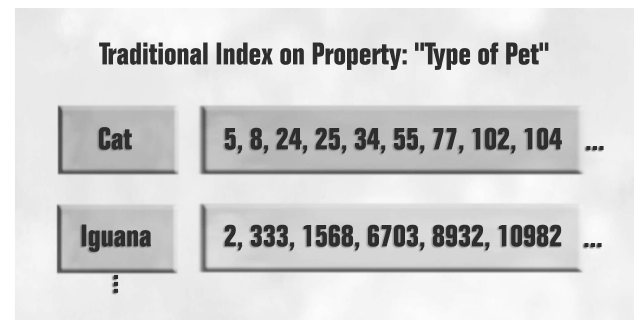
Caché supports shadow server technology. Its high performance minimizes the time required both for updating the shadow database, and for responding to queries.
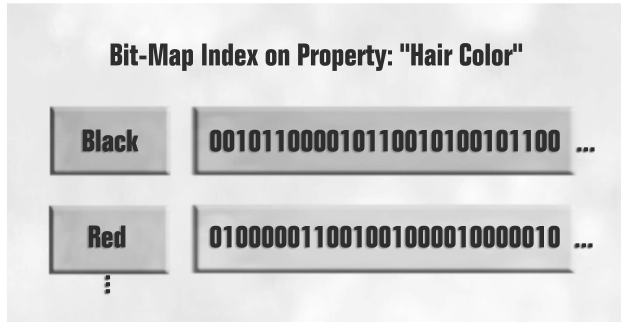
Caché for Real-time Data Analysis

Sometimes, even the few seconds of data latency that may occur in shadowed systems is too much. Sometimes it is necessary to run queries directly against rapidly changing operational data. In those cases it is vital that query processing be extremely fast, in order to avoid degrading the overall application performance.

In any data analysis system, one way to improve query response times is to create indexes on properties that are frequently used as criteria for searching the database. Caché gives developers the option of building both traditional and bit-map indexes.

In a traditional index, a property (in relational terms, a column of a table) is described by lists of record IDs. For each property value there is a list of record IDs for which the indexed property matches that value.

Traditional Index on Property: "Type of Pet"

| Cat | 5, 8, 24, 25, 34, 55, 77, 102, 104 ... |
| Iguana | 2, 333, 1568, 6703, 8932, 10982 ... |

Bit-map indexes are frequently used in data warehouse solutions because they can significantly speed up the analysis of large amounts of data. In a bit-map index, a property is described by strings of bits. For each property value there is a string of bits representing each record. The bit is "1" if the record has that property value, "0" if it does not.

Bit-Map Index on Property: "Hair Color"

Black	0010110000101100101001011001 ...
Red	0100000110010010000010000010 ...

The advantage of bit-map indexes is that queries can be processed by performing Boolean operations (AND, OR) on the indexes – efficiently determining exactly which records fit the query conditions, without searching through the entire database. For complex queries, bit-map indexes can decrease response times by more than a factor of 100.

However, bit-map indexes are difficult to update if existing data is changed or deleted frequently. That is why they have traditionally only been used in situations (such as data warehouse solutions) with static, read-only data. Bit-map indexes can also require large amounts of disk space, because every possible value in the index contains a bit (either 1 or 0) for every record.

Caché is the only database that enables "transactional bit-map indexing". Its efficient multidimensional data structures and sophisticated compression techniques combine to make Caché's bit-map indexes both smaller and much quicker to update than those offered by other database vendors. With Caché, bit-map indexes take no longer to update than their traditional counter-

parts, making them suitable for use with rapidly changing data.

Transactional bit-map indexing gives developers the capability to create "business intelligent applications" that can analyze rapidly changing operational data, and act upon the results in real time.

Ensemble for Business Activity Monitoring in Integrated Systems

The ability to analyze data in real time is also required by integrated systems that provide Business Activity Monitoring (BAM) capabilities. In the case of BAM solutions, the information being instantaneously collected and analyzed comes from a variety of disparate sources.

Ensemble, InterSystems' universal integration platform, has a built-in data repository where it stores every message that passes through an integrated system. This "message warehouse" uses the same multidimensional data structures as Caché, and provides the same lightning-fast performance. Ensemble can analyze data in its message warehouse in real time, which enables powerful debugging and management capabilities – and BAM.

Ensemble comes with features designed to speed the development of BAM solutions. Ensemble's advanced object technology and wizard-driven Studio make it easy to define "Business Metrics" that gather, or otherwise calculate, information about an integrated system at developer-defined intervals. That information can be associated with meters displayed on a customizable Web-based "executive dashboard". Business Metrics can also contain logic that implements a feedback loop within the integrated system. For example, when a measured parameter goes outside of its expected limits, the Business Metric can automatically take corrective action, provide notification, etc.

Conclusion

InterSystems' products, Caché (the high-performance database) and Ensemble (the universal integration platform), use a multidimensional data architecture that is well-suited to business intelligence solutions. Traditional data warehouse applications can benefit from the lightning-fast performance of Caché. But more importantly, such high performance enables real-time analysis of rapidly changing operational data, which opens the door for new kinds of business intelligence solutions. "Business intelligent applications" are standalone systems that can analyze and respond to information in real time, allowing enterprises to take full advantage of the data they collect as part of their everyday operations.

Ensemble enables the real-time analysis of messages and other data within integrated systems. With Ensemble, it is easy to build the executive dashboards and feedback mechanisms that are needed for Business Activity Monitoring solutions.

InterSystems Corporation

World Headquarters
One Memorial Drive
Cambridge, MA 02142-1356
Tel: +1.617.621.0600
Fax: +1.617.494.1631

www.InterSystems.com

INTERSYSTEMS