



Evaluating Real-Time Data Integration Solutions

Evaluating Real-Time Data Integration Solutions

Introduction

Access to timely and accurate information has become a critical driver of business success. Companies need to have access to timely accurate information to make better business decisions, serve their customers better and respond to changes in the market.

But increasingly, companies are hitting roadblocks in their quest for up-to-date accurate business information. Most mid- to large-sized corporations today have a data warehouse capable of analyzing information from production systems on a daily or weekly basis, however many are finding that extracting information from over-taxed mission-critical production databases is becoming more time consuming and complex. At the same time, the demand for more timely information is increasing to the point where daily or weekly updates are no longer sufficient.

This paper will examine the business issues driving organizations to consider real-time data integration as well as the strengths and weaknesses of different approaches to solving the problem. It will also provide an overview of DataMirror's solutions for real-time, low impact data integration.

The Demand for Real-Time Integration

What keeps today's Chief Information Officers (CIOs) up at night? It's likely a long list, and chances are data integration ranks high. According to a Forrester survey of more than 600 technology decision makers and enterprises across North America, improving integration between applications is one of the top priorities for organizations in 2006.

Why? Let's look at some of their business challenges:

1. **Increasing demand for real-time information for reporting and analytics.** Traditionally, reporting has been done from warehouses which are updated on a daily or weekly basis. For many types of reports that data is current enough. For others, though, nothing short of up-to-the-minute will do, such as inventory data where product inventory is very high or very low, or billing information where billing is done by the minute or fraction of a day.
2. **Large volumes of information are difficult to handle in a batch window.** As more information is gathered – such as online transaction data, inventory data, and customer information – the effort involved in moving it to the warehouse increases drastically. Many companies are finding that an eight-hour batch window is no longer sufficient for traditional ETL (extract, transform, load) tools to integrate all of the needed data.
3. **Necessity to conduct business 24/7 is reducing batch windows.** As more business is done across time zones and over the web, many companies are faced with the problem of shrinking batch windows, making it more difficult for traditional ETL tools to extract data in the short time available.

4. **Growing need to detect and react to business events as they happen.** Many companies are looking for ways to detect business events in production systems and have those events trigger a response in another system. For example, a cell phone company would like to send a text message to a customer running low on minutes asking if they would like to purchase more.
5. **The need to track all changes for auditing purposes.** Companies need to comply with regulations, which often require them to continuously track all changes to data, not just the net result of those changes.
6. **Increasing need to keep data in sync across the enterprise.** Customers want up-to-the-minute access to order, payment and inventory data so they can buy products, pay bills and check delivery status online. Employees need much of the same so they can better service customers and make wise business decisions. To accomplish this, eCommerce data needs to be in sync with business applications and data needs to flow in real-time across the enterprise.
7. **Companies want to deploy new applications using data on legacy systems without paying for an increase in workload.** Often, legacy systems are already maxed out and new capacity is very expensive. Companies want new applications on UNIX or Windows to avoid this cost, but integrating the data from those legacy systems without increasing the load on them is a key issue.

Why Traditional ETL Tools Aren't Always the Solution

The majority of corporations use some type of ETL tool to extract data in bulk from their production systems and load it into other systems including data warehouses. The main strengths of traditional ETL tools are that they extract data from many different applications, perform complex transformations on that data, and then bulk load large volumes of data into data warehouses.

But when it comes to extracting data from production systems, there are several limitations to how an ETL tool can be used. These limitations include:

1. **ETL is a batch process that requires a batch window** – ETL tools can move batches of data out of production systems into data warehouses. For many corporations, running a nightly batch process when systems are not running at full capacity works well. However, as businesses are increasingly globalizing and as more and more business is conducted 24/7 over the Internet, finding a time when production systems can be taken offline to run a batch extraction of data becomes increasingly difficult.
2. **ETL batch windows increase with the amount of data extracted** – Even where corporations can tolerate an overnight batch window, many are finding that with increasing amounts of data to be extracted the batch window required to do the extraction is expanding beyond the eight to 12 hours that their production systems are offline.
3. **ETL tools typically require that changes be made to the production systems they extract data from.** ETL tools often require changes to production systems – for example, adding a date/time stamp to allow incremental updates. For many corporations, the additional risk this would pose to their mission-critical systems is not something they can tolerate.
4. **ETL tools do not track all changes that have been made to the system. Rather, they just track the net result of those changes.** For example, if the inventory level of a certain product decreases due to purchases but then increases due to restocking, or if the status of an order changes, typical ETL tools will not track this.

Change Data Capture Technology and Real-Time Data Integration

As an alternative to traditional ETL technologies, Change Data Capture (CDC) technology provides a way to capture changes on production systems so that they can be applied elsewhere without directly querying the database.

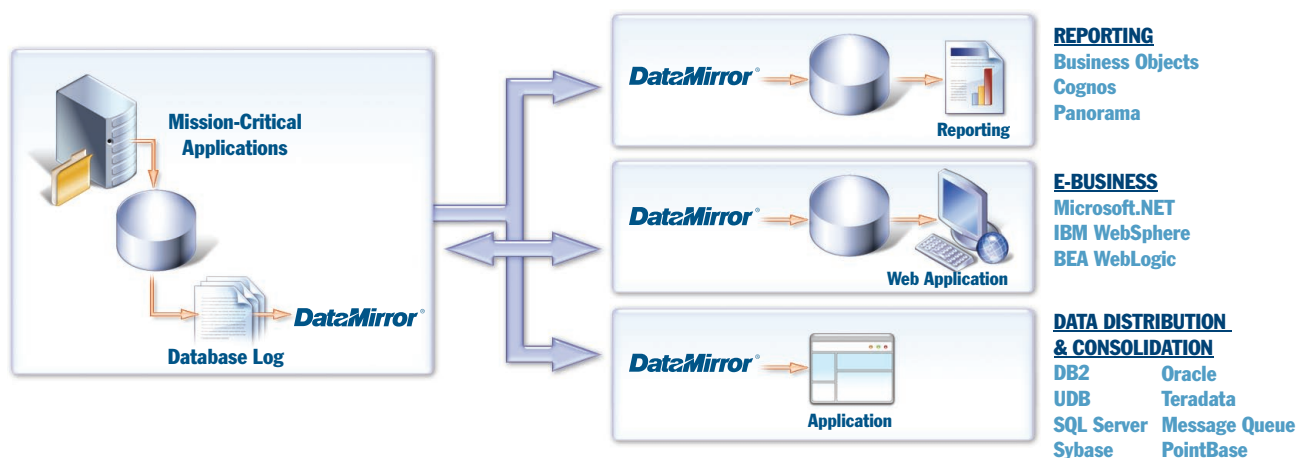
CDC has many benefits over traditional ETL tools and, in many cases, is complementary to traditional ETL tools. Benefits over ETL include:

- **Changes are captured in real-time so information is always up-to-date.** CDC captures changes continuously as they occur. The result: information is always up-to-date rather than being only as current as the last batch window.
- **No impact on the performance of production systems.** CDC reads database log files rather than querying the database directly. There's no impact on mission-critical production systems.
- **No requirement for batch windows.** With changes captured, transformed and applied continuously, there is no need to take the systems down to extract data.
- **Easily scales to very large databases and large numbers of transactions.** Only changes are replicated rather than all of the data in the changed tables. The result is much greater scalability through less data being moved across.
- **Does not require changes to the source system.** Because CDC is only reading the log, it does not require changes directly to the source database, yet it can detect all transactions including descriptive information about the change (user, application, time, etc).
- **Logs all changes to the system, not just the net results of those changes.** For audit and compliance, all insert, update and delete actions are recorded rather than just the net results of those actions.
- **Can complement ETL tools.** Many corporations combine the strengths of CDC and ETL tools. For example, real-time CDC technology accesses the data directly from database logs and replicates data from operational systems without requiring batch windows or interruptions to mission-critical systems. The data is then loaded into operational data stores (ODS), with the ETL tool then feeding the data warehouses or other applications. Production systems can be kept up and running 24/7 and real-time reporting can be supported off of the ODS.

About DataMirror Transformation Server

DataMirror's Transformation Server software provides real-time, bi-directional data integration and transformation between diverse relational databases and other data stores on different platforms.

Unlike traditional ETL, batch load, query or message-based systems, Transformation Server was designed from the very beginning to deliver scalable, high performance data integration in real time with minimal latency. Transformation Server's unique Change Data Capture technology has a minimal impact on the performance of operational applications.



Benefits of Transformation Server

- **Low Impact** – Rather than using triggers or performing queries against the database, Transformation Server uses best-of-breed log-based Change Data Capture (CDC) technology to capture changed data from database logs. This ensures that the performance of even the most demanding mission-critical applications running on the source system is not adversely impacted.
- **Easy to deploy** – Transformation Server's easy-to-use GUI makes it simple to select source and target databases and then configure transformations. The net result: faster implementation, lower cost, and increased return on investment.
- **Real-time** – Data changes are captured in the source system as they happen and the changes flow immediately to the target systems. All information is up to date and in sync.
- **Flexible** – Customers use Transformation Server to solve many different business problems across a wide range of platforms – from mainframe to mobile. Use it to load a data warehouse, synchronize data between existing systems and web applications, distribute and consolidate data between different applications, or manage other replication-based requirements. It can be deployed in a variety of architectures, on many different platforms, and between all major databases including DB2, UDB, Microsoft SQL Server, Oracle, Sybase, Teradata, and PointBase.
- **Scalable** – Transformation Server is a high performance solution that easily scales to large data volumes without impacting the performance of mission-critical applications running on the source system.

Who Needs Transformation Server?

- Companies that want to load their data warehouses in real time so they can make operational and tactical business decisions based on the latest information.
- Companies that want to provide customers with access to real-time inventory information from their website while minimizing the impact on their mission-critical logistics systems they use to ensure efficient continuous operations.
- Companies that have multiple production applications or have recently undergone a merger or acquisition and need to consolidate and distribute information between diverse systems.
- Public companies that need to conduct consolidated financial reporting across a range of transactional systems in different regions, departments, or business units.
- Businesses that need to improve the operational performance of their systems, which is being adversely affected by shrinking nightly batch windows and expensive query and reporting functions.

For more about Transformation Server, visit www.datamirror.com/products/tserver.

A Real-Time Solution for Real World Problems

Reporting and Analysis

Priority Health, a Michigan-based managed care company, uses Transformation Server to offload query and reporting activities from the production server and maximize operational efficiency. The company's entire transaction history is captured from the production Sybase database and flowed into an HP-based Oracle operational data store.

Transformation Server captures, transforms and flows hundreds of critical data transactions per second into the Oracle operational data store so that they can be stored and easily accessed for reporting and Web-enabled applications.

In the past, Priority Health used a combination of custom-developed batch solutions and various integration software packages to periodically move data across their multi-platform environment. But now, with Transformation Server, Priority Health has the capability to deliver absolutely current and up-to-the-minute information for business intelligence and real-time analytics. Powerful data transformation on-the-fly ensures that the operational data store is populated with clean data in a format conducive to query and analysis.

“Because of Transformation Server’s ability to deliver real-time data from Sybase to Oracle, we anticipate that approximately 250 in-house employees and decision makers will access the replicated data via various query and analysis tools, while many others will access the data via the web,” says Keith Sederholm, Manager, Database Services, Priority Health. “By implementing Transformation Server, Priority Health’s diverse technologies can coexist while the data is kept current, synchronized and secure.”

To read the full story, visit the DataMirror Customers page at www.datamirror.com/customers.

Supporting eBusiness

Acushnet Company, manufacturers of leading golf brands including Titleist, FootJoy, Cobra and Pinnacle, uses Transformation Server to bi-directionally integrate its production data between its iSeries-based order management system and a SQL Server database running on Microsoft Windows, which supports its customer extranet. This offers customers and employees real-time access to order, payment, credit and inventory data. The Company's extranet handles business-to-business transactions for up to 14,000 customers.

"We wanted to initiate a real-time e-Business extranet on a Microsoft SQL Server database that would empower customers through up-to-the-minute access of our production data," says Mark Koerner, Systems Integration Specialist for Acushnet. "The problem was that we did not want to build the extranet to access our iSeries legacy data directly. We needed an efficient solution with filtering capabilities that would seamlessly integrate our back-end legacy systems with our new e-Business environment."

Acushnet is also using Transformation Server to integrate data from its production iSeries database to another SQL Server database used for reporting purposes by hundreds of employees throughout the enterprise.

Adds Peg Nicholson, SVP and CIO of Acushnet: "Here we have a single, flexible product that can simultaneously support our e-Business and data warehousing efforts, as well as support any new business applications that we may later envision."

To read the full story, visit the DataMirror Customers page at www.datamirror.com/customers.

[About DataMirror]

DataMirror (TSX: DMC), a leading provider of real-time data integration, protection and Java database solutions, improves the integrity and reliability of information across all of the systems that create and store data. DataMirror's flexible and affordable integration solutions allow customers to easily and continuously detect, translate, and communicate all information changes throughout the enterprise. DataMirror helps customers make better decisions by providing access to the continuous, accurate information they need to take timely action and move forward faster.

Over 2,000 companies have gained tangible competitive advantage from DataMirror software. DataMirror is headquartered in Markham, Canada, and has offices around the globe. For more information, visit www.datamirror.com.

Copyright © 2006 DataMirror Corporation. All rights reserved. The DataMirror family of related marks, images, and symbols are the exclusive properties and trade-marks of DataMirror Corporation-used under licence. All other brand or product names are trademarks or registered trademarks of their respective companies.

How to reach us:

North America 1 800 362 5955
UK + 44 (0)20 7633 5200
France + 33 (0)1 72 75 73 40
Germany + 49 6151 8275 0