



GoldenGate For Real-Time Data Warehousing

Table of Contents

Introduction	3
Doing Business in Real Time	3
Meeting The Challenge –	4
Data Warehousing Yesterday & Today.....	4
Staying Ahead of the Curve	5
The Need for Data Movement with Speed, Volume & Diversity	6
Making it Work in the Real World.....	6
Westcon Oracle-to-SQL Server Data Warehousing for Improved Performance.....	6
Montefiore Medical Center / EHIT Real-Time Outcomes Analysis for Cost Reduction.....	8
Sabre Holdings, Inc. Open Systems Synchronization for Low-Price Air Fare Look Up	9
Speed, Volume & Diversity = Performance	11
The Critical Success Factors For Implementing Real-Time Data Warehouses	11
What to Look For.....	13
Performance.....	13
Usability.....	14
Industrial Strength	15
Diversity.....	16
Conclusion	16

Introduction

It is the purpose of this paper to detail how GoldenGate Data Synchronization can deliver up-to-the second data to support real-time reporting and analytical applications. Data warehouses have long been used to compile and generate critical information, about the business, and for the business. As the need for more timely information has grown steadily in recent years, however, so has demand risen on reporting infrastructure. Where ten years ago an organization would have found it adequate to perform online reporting off a production system once a week, it may now require continuous offline reporting throughout the day on multiple aspects of the business. In other words, real-time data warehousing. To illustrate the role GoldenGate can play in providing a real-time data warehouse capability, this paper will provide:

- a brief overview of business and competitive drivers behind the growing need for real-time data
- a retrospective look at data warehousing techniques and methods
- examples of successful implementations demonstrating various architectures for providing access to real-time data
- the critical technology evaluation criteria when real-time data warehousing is a requirement

Doing Business in Real Time

Of all the changes wrought by the Internet over the past decade, one of the most significant is the acceleration in the pace of business. The phrase “real-time” has become something of a cliché in recent years, but for many organizations, getting information and services to decision-makers, partners and customers when they want it and where they want it has simply become a reality of doing business. And for the IT managers tasked with managing enterprise data, these new business drivers directly translate into a need to rethink traditional approaches to systems architecture.

As individuals have come to expect instantaneous online shopping, banking and access to information, businesses have had to respond by devoting more and more resources to meet the growing demand. In some cases IT departments have had to synchronize changing data from high volume transactional databases to lower-cost databases in order to support secure, web-based applications. Likewise, as constituencies within the organization have demanded more immediate access to information on sales metrics, supply chain, operations and financials, a growing number of IT departments have had to find innovative ways of gathering this information from various points in the organization, and then delivering it wherever needed. In both cases, a logical and effective way to realize these goals has been to implement active, or real-time data warehouses.

Meeting The Challenge – Data Warehousing Yesterday & Today

The 1990s saw the rise of the enterprise application as ERP and CRM technologies enabled executives and decision-makers to generate real-time reports and analytics on a wide range of measures, from financials to sales to general customer and marketing metrics. And as enterprise applications gained popularity, so grew the need for the data warehouse. There were several reasons behind this trend; first was system performance and the limitations of the installed infrastructure. In many implementations enterprise applications were hosted on transaction servers and pulled data from transaction databases, in other words: online reporting. As the organization grew—and the number of people updating the database and accessing the applications expanded—system overhead increased and performance steadily degraded. A major factor at play was that users who constantly update databases require quick execution of transactions, while users accessing reports often need to execute large queries. As the two needs directly compete for processing resources, it was not uncommon to find systems being brought to their knees because of an ill-timed large query.

To accommodate a large increase in users, organizations were confronted with the need to purchase bigger systems; meaning increases in memory, disk space and CPU. Despite the significant drop in hardware costs and increases in processing power over the years, most organizations still found the cost of additional transaction production infrastructure to be prohibitive. Some organizations turned to mixing transaction processing with ad-hoc queries. This method, however, also causes contention. Unscheduled queries can adversely affect online transactions when several queries are initiated at the same time, causing spikes in performance, destroying response times and bringing about shortfalls in service level commitments. In many cases a database fine-tuned for online transaction processing may not be conducive for executing ad-hoc queries. For example, an index may be created to provide fast query execution, however the penalty that must be paid is that the index itself may require update and maintenance as transactions are added, removed or updated from the database. The inevitable result is that processing is slowed down.

For these reasons, many organizations implemented data warehouses comprising separate servers and databases. Data from the production databases could then be copied to the data warehouse where the reporting applications could perform queries without negatively affecting the transaction processing system. An additional benefit was that data warehouses ensured superior operations on the reporting applications as well, because users had an environment and database fine-tuned for their needs.

Staying Ahead of the Curve

Problem solved? Not completely. Moving data from the production database to the data warehouse can typically be accomplished using extraction, transformation and load (ETL) utilities. These tools are reliable and capable of performing many of the tasks a data warehouse needs. However, nearly all ETL tools are batch-oriented, which means there is an inherent lag between the time data changes on the production system and when it is available for query by business users. In a typical implementation, an ETL utility can provide “snapshot” data movement once a week, once a day, or if data volumes are small enough, several times during the day. In the pre-Internet era, the inherent lag time incurred by this kind of batch processing was perceived as acceptable. Today, however, as the need for more up-to-date data has become more acute, many organizations are now looking for alternative solutions that can populate data warehouses in real-time.

Another problem with ETL tools is that they update systems with complete files, even if only small portions of the file in question had changed since the last batch update. This inefficient process bogs down servers and bandwidth, straining the systems to work harder than necessary and degrading overall performance. As data loads grow ever larger, this practice of loading complete files becomes a major problem for many organizations since the time window to populate a data warehouse grows proportionately with the increase in data and transaction volumes. In short, ETL tools may not be just a limiting factor to the “freshness” of the data, but in cases of extreme volumes they may also be unable to keep up.

ETL tools do work well pulling data from a variety of sources and transforming that diverse data for loading into a data warehouse. But again, where data volumes are very high, they can only perform at a relatively slow rate because they update complete files. So, if the batch window for loading a large database from several sources is once a week, ETL works well. Ask an ETL tool to replenish a large data warehouse several times a day, however, and the batch window will simply not be able to accommodate the data volume

ETL isn't the only approach to populating data warehouse. Many organizations leverage the expertise of DBAs or programmers to complement ETL tools using custom-coded scripts, but this can be a costly layer to add to an infrastructure. Specialized coding also delays implementation of new features that users require, which can reduce system flexibility and decrease access to timely data. In addition, scripts are often manually done and hard to maintain as the infrastructure changes. For instance, if a database is upgraded to a new version, the scripts may need to be re-programmed. Another method that has gained adherents in some sectors is message-based middleware, which has proven effective for integrating data in real-time. But middleware products cannot handle the large volumes of data that are necessary for running enterprise applications.

The Need for Data Movement with Speed, Volume & Diversity

To sum up, as the pace of business accelerated through the 1990s, traditional data warehouse “best practices” were hindering overall performance—just as online reporting practices had become obsolete years earlier. For more and more companies, pressure was growing to increase the capabilities and timeliness of the data warehouse. Which meant finding a solution that would allow them to:

- Increase the speed of data movement toward the goal of real-time
- Handle the volumes of data streaming between production systems and data warehouses
- Allow changed data from multiple sources to be captured and delivered into consolidated data warehouses providing a single unified view of the enterprise

Making it Work in the Real World

The following three real world examples illustrate how some businesses are successfully implementing real-time data warehousing with the help of GoldenGate:

Westcon

Oracle-to-SQL Server Data Warehousing for Improved Performance

Westcon is a leading global channel provider of networking products, Internet access, and security equipment, meeting the diverse needs of solution providers, resellers and system integrators. The company’s ERP application, JD Edwards OneWorld XE, had long shared space on an Oracle database server that was also used for transaction processing. As Westcon’s business grew, its reporting needs increased to the point where running queries on Oracle took up 95% of the CPU. With responses slowing to a crawl and transaction performance suffering, the company decided to implement a new Microsoft SQL Server data warehouse, which would function as a dedicated report platform. In this way, the reporting team would have uninterrupted access to the necessary data at all times, while the transaction server could have the total system resource to handle real-time transactions.

Initial tests with the new system were satisfactory, but once the full project was underway it became clear that software used to capture and move data from the transaction system to the data warehouse was not adequate. Due to the fact that the vendor used a trigger-based method for the Oracle to Microsoft SQL Server replication, it soon became apparent that the solution was not effectively scalable for Westcon’s enterprise application, which uses more than 1,500 tables. This

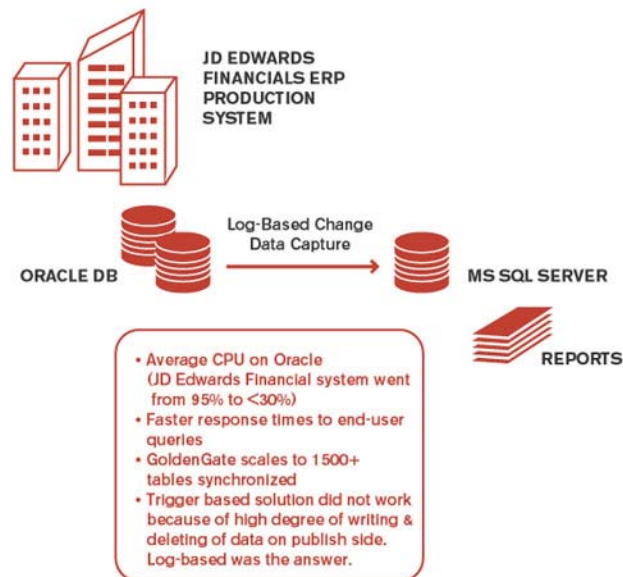
situation led to two problems: First, the load on the database was higher than expected because of the constant writing and deleting of data to log tables on the publisher side. Second, and most importantly was the fact that should any issue occur with the log table then any table that was being replicated became inaccessible as the database triggers could not fire. This meant there was a potential for system wide data integrity issues as well as production downtime.

To remedy the situation, Westcon put GoldenGate's Data Synchronization solution into production to continually move data between its Oracle and SQL Server databases. Since GoldenGate uses log-based data capture, it consumes far less system resources on the publish side than trigger-based solutions. With its new off-line reporting infrastructure in place Westcon saw CPU usage drop to about 30%. Moreover, since GoldenGate can scale to meet the data volume needs of even the largest global enterprise, Westcon can rest assured that its synchronization capability will more than adequate well into the future.

"The GoldenGate services and support team did an outstanding job. Today, GoldenGate is in production at Westcon and the performance is exceptional, as advertised, sold and delivered."

Stuart Morrison
Westcon

Westcon: Real-time Reporting for Business Users



Montefiore Medical Center / EHIT

Real-Time Outcomes Analysis for Cost Reduction

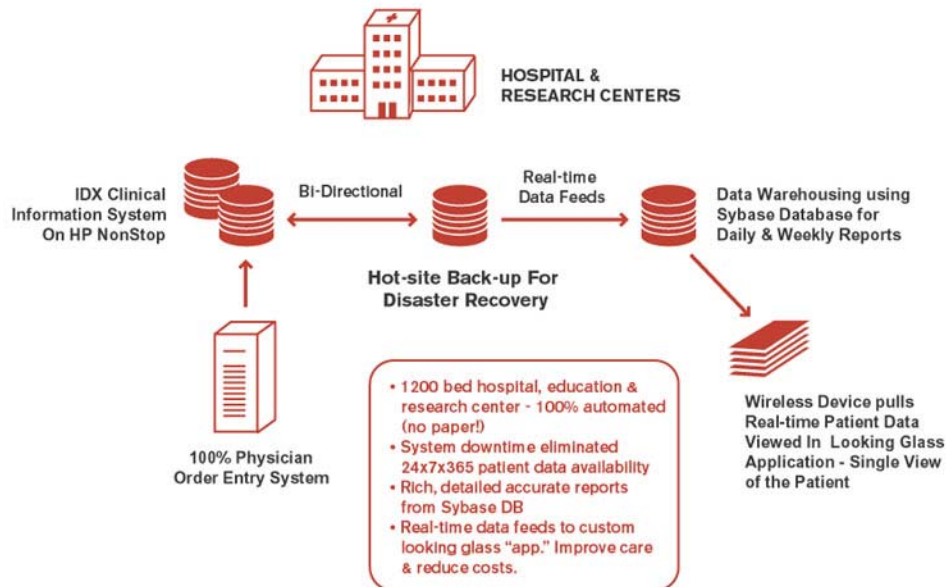
Like all healthcare providers today, Montefiore Medical Center is under enormous pressure to reduce costs while maintaining high standards of care. In response, Montefiore, working with its wholly owned subsidiary Emerging Health Information Technologies (EHIT), has implemented several IT systems to automate care, eliminate errors and streamline decision-making. The focal point of this effort is the hospital's physician order entry/clinical information system (POE/CIS) provided by IDX Healthcare. A transaction-based production system, Montefiore's POE/CIS is a powerful tool for optimizing clinical care because it collects information from sources all across the hospital—laboratory, pharmacy, admitting, treatment rooms, and so on. In this way, it provides immediate access to all the important data doctors and nurses need to make responsible clinical decisions. Moreover, Montefiore's executives realized that as the repository for detailed information on clinical orders—including prescriptions, tests and therapies—the POE/CIS could provide insight into treatment practices and outcomes.

To tap into this vast data store, EHIT and Montefiore designed and implemented a data warehouse using GoldenGate's data synchronization software to capture, transform and deliver massive volumes of data at sub-second speed from the hospital's POE/CIS to a Sybase database. With GoldenGate in between, this real-time data warehouse could pull data from multiple sources across the hospital, perform transformations on the fly, and make the data available for Montefiore's Clinical Looking Glass application, which was developed in house. Using this application, doctors and technicians at Montefiore were able to perform real-time decision support and outcomes analysis specific to patient care. With this longitudinal evaluation capability, the hospital can proactively monitor the quality of care down to the individual level, and begin implementing targeted care improvement efforts that help contain costs, eliminate errors and improve clinical decision-making.

"GoldenGate allows us to synchronize the data between both systems, allowing us to benefit from the best of both worlds. We manage care and analysis, GoldenGate manages the synchronization."

*Dr. Eran Bellin
Director of Medical Informatics
EHIT*

**Montefiore Medical Center/EHIT:
100% Automated and Real-time Decision Support**



Sabre Holdings, Inc.

Open Systems Synchronization for Low-Price Air Fare Look Up

When Sabre set out in 2001 to build the Air Travel Shopping Engine (ATSE), a new low-price airfare search and itinerary service, the company's IT brain trust decided to diverge somewhat from its traditional technology approach. Sabre Holdings, of course, is renowned for its airline reservation network, which was the world's first computerized transaction-processing system, and for Travelocity, the groundbreaking website that set standards for usability and performance in the mid-1990s that were widely influential in the emerging e-commerce space. Driven by the growth of the Internet, the dynamics between these two mainstays of the Sabre Holdings business have changed substantially in recent years.

The company's IT infrastructure was originally designed for the needs of professional travel agents who used the Sabre system to perform quick, tailored itinerary searches and purchase transactions. With the growing popularity of the Travelocity site, however, the system now had to accommodate millions of individual travel shoppers who prefer to browse around comparing flights, airports and itineraries. This kind of complex, high volume fare searching is very taxing on IT resources and doesn't necessarily result in bookings. Instead of

serving a relatively small customer base of “bookers,” Sabre now had an enormous customer base of “lookers.”

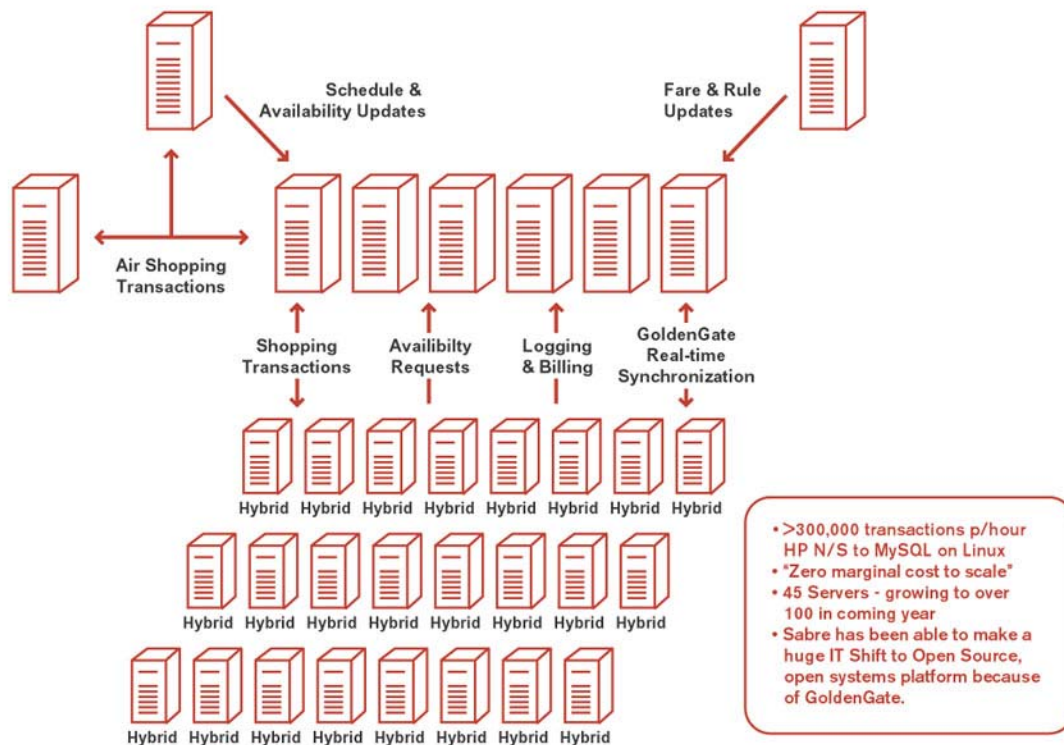
With the ATSE project, Sabre Holdings saw an opportunity to implement a secondary, non-production platform using low-cost open source technologies that would accommodate its shifting business model. In effect, Sabre would use GoldenGate synchronization to move data in real time from its core HP NonStop production systems over to a Unix- and Linux-based server farm running MySQL databases and the company’s proprietary ATSE application. By populating the data warehouse with real time information on flights, seating and prices from the NonStop production system, Sabre enabled its customers to perform their own searches and find the exact flight and price that best fits their buying criteria.

This innovative design solved a host of problems for Sabre. First, since all the consumer traffic would be handled on the open-system ATSE, it eliminated the drain on the HP NonStop production systems. As consumer traffic rose, Sabre could quickly and inexpensively scale up processing power by adding more commodity boxes to the server farm rather than the far more expensive HP NonStop system. And most importantly, with this active data warehouse Sabre was able to meet the needs of its changing customer base.

"GoldenGate has been a valuable intellectual partner in our ATSE platform implementation. Together we are honing Sabre Holdings' competitive edge as we advance our travel commerce technology."

Craig Murphy
Chief Technology Officer
Sabre Holdings

Sabre Holdings: Driving High-speed Searches for Travelocity



Speed, Volume & Diversity = Performance

The Critical Success Factors For Implementing Real-Time Data Warehouses

The three examples above illustrate that while organizations may have different business requirements, they all share the need for speed, volume and diversity. In fact, what these case studies show is that the only way to effectively perform real-time data warehousing in an efficient and cost effective manner is to do it with a data synchronization tool. Unlike ETL tools, GoldenGate data synchronization uses change data capture technology, a highly efficient, non-intrusive means for capturing thousands of database operations per second. Because it only replicates data that has changed on the transaction production system, change data capture synchronization results in an ultra-light footprint on existing systems while offering outstanding reliability and fast implementation.

Speed

In addition to enabling the throughput of thousands of transactions per second, data synchronization incorporates design features and attributes such as automatic filtering, transformation and mapping, smart transaction grouping and native database access to ensure fast processing. What's more, parallel processing architecture and compression technology is leveraged to provide high performance and high-speed throughput using a minimum of network bandwidth. Finally, transactions are delivered in logical sets to ensure data integrity at all times.

Volume

Data synchronization using change data capture technology is a very efficient way to handle massive amounts of data quickly and efficiently. It executes multi-record transfers of data to and from disk as well as over the network when capturing, transforming and delivering database changes. Moreover, it leverages buffering techniques to balance throughput and latency considerations. For instance, processes can be posted to run local to the target database, maximizing throughput by avoiding network limitations. The platform can also group smaller transactions into larger commit groups to further minimize disk IO while preserving original transaction properties. Finally, data synchronization is designed to execute updates via native database interfaces rather than through middleware, and utilizes internal caches to ensure fast execution of repetitive statements.

Diversity

It is impossible to execute true data synchronization organization-wide and on a global scale unless you have a platform that is compatible with all major operating systems and database environments. GoldenGate offers this capability, working with most of the popular databases today, including Oracle, DB2, Microsoft SQL Server, Sybase, Informix, NonStop SQL, Enscribe, Teradata and TimesTen, as well as up-and-coming databases such as MySQL. Moreover, GoldenGate leverages native interfaces to eliminate the need for additional middleware when capturing, transforming and mapping data between these diverse databases. Finally, GoldenGate provides further coverage through ODBC to ensure that all other possible combinations are covered. When it comes to enterprise databases, hardware and operating systems, no other synchronization software works across such a wide range of systems and databases:

Databases

Oracle
DB2
Microsoft SQL Server
Sybase
Informix
Enscribe
All ODBC compatible databases

**Operating Systems/Hardware
Platforms**

Unix
Windows NT / 2000 / XP
Linux
HP NonStop
Sun Solaris
HP / UX
IBM AIX
DEC VMS
IBM OS/390
SCO
Siemens
TRU64

What to Look For

Since GoldenGate first introduced its synchronization software in the mid 1990s, several other change data capture offerings have become available. Only GoldenGate, however, delivers the three key criteria of speed, volume and diversity for the full range of enterprise infrastructures. Beyond these key factors, IT executives who are considering data synchronization tools for a real-time data warehouse implementation, or upgrade of current data warehousing capabilities, must examine attributes such as performance, usability and other features. Among the essential questions to ask:

Performance

TECHNICAL FUNCTIONALITY	CONSIDERATIONS
Scalability	<ul style="list-style-type: none">- Have you projected your highest levels of volumes and will the technology sustain adequate performance?- Can each functional area of the technology scale: Source Capture, Data Transformation, and Target Delivery?
Native Log Capture	<ul style="list-style-type: none">- Does the technology monitor and capture data changes with a native log based capture that has proven to be the preferred approach?- Does each data environment you require use the same log capture technique?- Does the technology support multiple log capture processes from the same data source?

Transaction Grouping	<ul style="list-style-type: none"> - Do the data capture and delivery approaches support logical transaction groupings of the data for both the source and target to insure effective data delivery of data to business intelligence and data warehousing applications as well as your ability to optimize data recovery and rollback if required?
Architectural Design	<ul style="list-style-type: none"> - Is the technology designed to provide unprecedented speed? - What key attributes deliver that speed? - Does the architecture allow for both speed and the ability to scale to the projected future volumes? - Is low overhead on the application process a design consideration?
Benchmarking	<ul style="list-style-type: none"> - Are there benchmarks to substantiate the vendor's claims? - Is the product being utilized in an environment that exceeds your expected volumes? - Have you done a "Proof of Concept"?

Usability

Transformation	<ul style="list-style-type: none"> - Can the technology be used to transform data between data environments, including same vendor versions and different vendor versions? - Is the source capture function effective and efficient without knowing where or what the target is?
System Integration	<ul style="list-style-type: none"> - Can the technology closely integrate with existing ETL and messaging products already installed?
Filtering	<ul style="list-style-type: none"> - How easily can selection criteria be applied to filter information? - Are stored procedures supported? - Is PL SQL supported? - Is Wildcarding supported?
Topologies	<ul style="list-style-type: none"> - To optimize your data architecture, are multiple topologies such as one to one, one to many, many to one, many to many and cascade architecture supported?
Flexible Peer / Peer Architecture	<ul style="list-style-type: none"> - Can the source and target data environments be on different platforms with different release versions AND can additional sources and targets be added WITHOUT impacting the existing data synchronization processing?
Support Array of Interfaces	<ul style="list-style-type: none"> - Are other queuing systems, XML, messaging and flat file output available? - Can the technology "integrate" into my existing EAI, ETL, and EII environments?

Alerting Mechanisms	<ul style="list-style-type: none"> - Can you monitor latency and architecture while data synchronization is active? - Can the technology facilitate notifications of alerts or launch specific actions? - Can you set up “alerts” to monitor and communicate defined situations, such as greater than 20ms latency between data environments and launch a variety of communication techniques?
No Downtime	<ul style="list-style-type: none"> - Is system downtime required to implement the technology? - Do you have to take any downtime for any upgrade of the technology?

Industrial Strength

Security	- Does the product have an encryption feature to protect the transmission of sensitive data inside and outside the business?
Compression	- In order to preserve precious network resources, does the technology leverage compression of the data that flows across the network?
Low System Overhead	- Does the overhead for the product not exceed a single digit systems overhead?
Resiliency	- Can the product sustain network or system outages without losing data?
Transactional Integrity	- Is transactional integrity maintained between source and target systems?
Guaranteed Delivery	<ul style="list-style-type: none"> - Are transactions delivered once and only once in the proper sequence? - What techniques are included to ensure guaranteed delivery?
Web Based Management	- Can the system be monitored and managed via a web-based interface?
LAN/WAN Support	- Are various networking topologies supported?
Use of the Network	- Given the precious nature of network bandwidth, are the data transmissions and connectivity messaging done in a responsible fashion?
Ease of Setup	<ul style="list-style-type: none"> - How easy is the initial set-up? - Is there downtime or extensive “programming” involved in set-up?
Ease of Administration	<ul style="list-style-type: none"> - Can the system be administered via web interface and automated scripts? - Is the administration managed by a DBA and is it intuitive?

Diversity

Supported Environments	<ul style="list-style-type: none">- Are each of the data environments that you require already supported?- Do you need to plan for new data environments from potential mergers or acquisitions?- Would you like to synchronize data with a customer or partner?- Are there legacy data environments that you would like to synchronize with your emerging data environments?- Are there data warehousing or business intelligence applications that would benefit from lower latency data synchronization?
-------------------------------	---

Conclusion

For many industries today, real-time has become the speed at which business is conducted. Decision-makers, customers, partners and suppliers have all come to expect access to up-to-the-second information, and IT managers have had to determine the most effective means of delivering this information. As shown above, active data warehousing using data synchronization is a proven means of achieving real-time capabilities for a wide variety of business needs, including internal reporting, outcomes analysis and customer service. More to the point, data synchronization has become an essential tool for managing the business, reducing the cost of doing business and adapting the IT infrastructure to meet the changing needs of the business.

Where real-time capability is a critical competitive business driver, data synchronization from GoldenGate is the answer. GoldenGate data synchronization software has been successfully deployed in thousands of implementations by many of the world's leading banks, financial institutions, retailers, media companies and healthcare organizations. It is the "One Investment One Platform" solution for your entire enterprise, providing both technical adaptability offering speed, volume and diversity, as well as a low total cost of ownership, ultimately offering you the key to doing business in real time.

For more information please visit our website www.GoldenGate.com