GOLDENGATE™

GoldenGate Global Data Synchronization Platform Technical Overview

A GoldenGate Whitepaper
January 2003

This paper presents an overview of the features and functionality of the GoldenGate Global Data Synchronization platform, and touches on many of the benefits it provides to the data-driven enterprise. The GoldenGate solution suite is continually upgraded to meet the changing needs of businesses, and will be developed further over time. Please refer to the GoldenGate Web site (www.goldengate.com) for the latest product and service offerings.

The Global Data Synchronization Solution

GoldenGate delivers Global Data Synchronization—the immediate capture, transformation and delivery of large volumes of data across your entire organization and beyond. For the real-time enterprise, GoldenGate provides the essential technology and expertise to implement a wide range of business and IT solutions, from disaster recovery and business continuity to application integration, system migration and active data warehousing. Where instant access and synchronization of up-to-the-second business information is essential, GoldenGate is the platform.

GoldenGate Capabilities

SPEED

Organizations implementing zero-latency business initiatives to improve and advance their decision-making and customer service capabilities must be able to capture, transform and deliver data at a rapid rate. The GoldenGate platform relies on proprietary change data capture technology that enables the throughput of thousands of transactions per second.

Change Data Capture from Transaction Logs

In the case of most databases, GoldenGate captures changes from the transaction logs, which are automatically maintained by database software and contain all inserts, updates and deletes performed by applications. Because GoldenGate captures data incrementally from this existing source there is no need to devote system resources to creating a secondary log of data changes. This method optimizes processing speed by minimizing overhead—typical GoldenGate implementations take up less than five percent of system resources—even in very high-volume situations. The result is the highest possible performance and scalability.

Trigger-Based Data Capture

Where deemed appropriate, the GoldenGate platform can also perform data capture using database triggers. In addition, GoldenGate can be configured to enable customer applications to create their own change sources for data capture, transformation and delivery.

Snapshot Data Capture

For purposes of initial synchronization, GoldenGate can replicate point-in-time table snapshots, either directly or by preparing input to high-speed loaders (such as SQL*Loader, BCP and DTS).

Designed for Fast Performance
GoldenGate incorporates numerous design features and attributes that ensure speedy processing, such as automatic filtering, transformation and mapping, smart transaction grouping and native database access. In addition, GoldenGate's parallel processing architecture and compression technology ensure high performance and speedy throughput using a minimum of network bandwidth. Whether performing simple synchronization or more complex transformation between dissimilar databases, GoldenGate carries out all required tasks with minimal configuration and no programming, enabling solutions to be deployed as quickly as possible.

GoldenGate in Action

Cable Television Service Provider–Disaster Recover
Within ten minutes of discovering a programming error, a major cable TV service provider recovered mistakenly deleted data by using GoldenGate to capture the deleted operations for the file in question and replicate those operations back as inserts.

Electronic Payment Solutions Provider–Real-Time Application Integration
The merchant clients of one of the world's leading electronic payment systems have access to real-time customer account information via Web-based reporting and analysis tools because the GoldenGate platform constantly feeds these applications with up-the-second data.

VOLUME
To support active data warehousing and business continuity systems, large organizations must handle data loads that reach into the tens of millions of transactions per day. The GoldenGate platform is specifically designed to handle massive amounts of data quickly and efficiently.

High-Performance Capture
GoldenGate executes multi-record transfers of data to and from disk, as well as over the network, when capturing, transforming and delivering database changes. In addition, advanced buffering techniques effectively balance throughput and latency considerations.

High-Volume Capture, Transformation and Delivery
GoldenGate posting processes run local to the target database, maximizing throughput by avoiding network limitations. The platform also groups smaller transactions into larger commit groups, minimizing disk IO while preserving original transaction properties. In addition, updates are executed via native database interfaces rather than through middleware, and internal caches are utilized to ensure fast execution of repetitive statements.

Efficient Network Utilization
The GoldenGate platform, which supports data delivery over WAN, LAN, Internet and fibre channel, reduces network bandwidth requirements in a number of ways. For update records, only primary key columns and columns that changed are transmitted, rather than the entire record. In addition, transaction log overhead, such as index records and before images, is filtered at the source. Traffic is optimized by packaging individual records into larger messages, and avoiding record-at-a-time bottlenecks. In addition, it provides several levels of data compression to further reduce the amount of network bandwidth required for transmission. In typical applications, data compression can be specified to reduce byte transfer by 75% or more.

GoldenGate in Action

Food Distributor–Business Continuity
One of America's largest restaurant and food service suppliers uses GoldenGate to synchronize data between its product database and its customer-facing e-business applications. With near-real-time data synchronization, the company ensures 100% uptime and keeps prices for its more than 135,000 stock-keeping units up to date, 24 hours a day, seven days a week.

ATM Network Operator–Business Continuity
The owner/operator of the busiest shared ATM network in the world—comprising more than 8,000 machines processing as many as 400,000 transactions per day—uses GoldenGate to power its hot-site disaster recovery system, and ensure business continuity around the clock.

DIVERSITY
IT environments within most large organizations today are made up of a wide range of applications, databases, operating systems and hardware platforms. Synchronizing data across these heterogeneous systems has long presented significant hurdles. GoldenGate provides a solution to these problems by enabling data capture, transformation and delivery between virtually all major database, hardware and operating system platforms.

Comprehensive Platform Support
GoldenGate can replicate data bi-directionally between any combination of Oracle, DB2, Microsoft SQL Server, Sybase, Informix, NonStop SQL and Enscribe databases, running on most popular operating systems. In addition, GoldenGate replicates data from any data source to Teradata and TimesTen, as well as to any ODBC-compatible database including, Progress and Access.

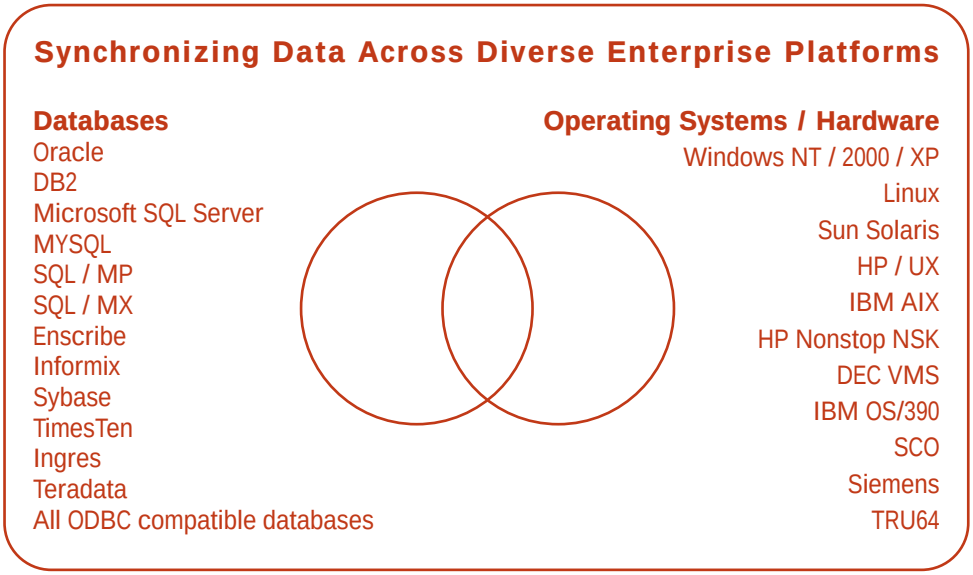


Figure 1 - GoldenGate Compatibility Matrix

No Intermediate Platforms or Databases Required

GoldenGate completes all processing tasks on the source and target platforms, without requiring an intermediate system in the middle. In addition, native database interfaces are supported for most databases, avoiding the need for additional middleware. Database changes are staged in queue files, reducing overhead and manual maintenance tasks.

GoldenGate in Action

Medical Center / Hospital Campus–Real-Time Reporting

Medical researchers, physicians and business employees of a major New York-based teaching hospital can now perform ad-hoc reporting and research based on real-time data, because GoldenGate allows for the sub-second capture, transformation and delivery of patient information between its HP-based physician order entry system and its Sybase data warehouse.

GoldenGate Components

Processing Overview

The GoldenGate Global Data Synchronization platform is made up of three distinct modules: Capture, Deliver and Manage. Capture and Deliver perform most processing functions, and these two modules are decoupled in order to provide maximum processing flexibility, modularity and performance.

Manage, the third major component of the platform, enables command execution and component management tasks. GoldenGate's basic processing model includes three main features:

- Database changes are recorded, either transparently by database software or via a GoldenGate capture mechanism
- Capture retrieves, filters, processes and writes selected data to a queue on the target system
- Deliver retrieves, filters and processes queued changes and posts updates to the target database

Note that GoldenGate customers often implement variations on this basic model in order to meet their own requirements.

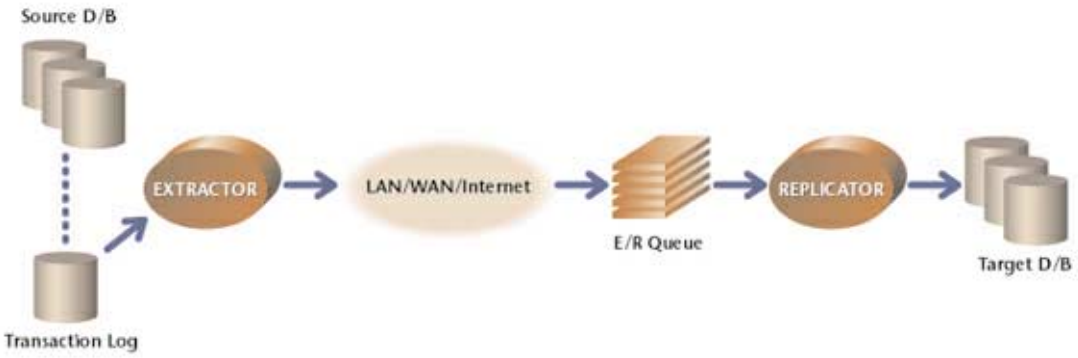


Figure 3 - Logical Architecture

CAPTURE

The core component for capturing data from source systems, GoldenGate's Capture module provides a host of flexible functions for queuing, formatting and transforming data. The Capture module:

- Retrieves inserts, updates and deletes from a database change source
- Filters out unnecessary data
- Selects specific columns and transforms the data (optionally)
- Outputs changes to one or more queues

Database Change Sources

To maintain optimal performance, Capture employs a range of database change capture techniques—each one optimized for the particular database. For Oracle, RAC, DB2, NonStop SQL, Sybase and Enscribe, Capture uses database transaction logs as its change source. Transaction logs contain all changes to the database and are automatically maintained by the database software, independent of Capture.

For Microsoft SQL Server and Informix, GoldenGate creates an additional log table within the database to capture changes. Change records are automatically created via Capture-generated database triggers and are stored in the log table. This log table acts as the source of changes for all tables in the application, or for selected tables. Capture automatically deletes changes from the log table after they have been processed.

For IMS, VSAM and non-audited Enscribe databases, changes are captured via operating system exits and staged into high-speed queues.

Capture can also capture:

- Point-in-time snapshots directly from database tables
- Changes logged by customer applications
- Change data from GoldenGate queues (for unique staging and cascading requirements)

Database Change Retrieval

Capture retrieves inserts, updates, deletes and transaction information from the database change source. As data is generated by the source application, Capture blocks and transmits data to minimize latency. It also periodically checkpoints its position in the change source to guarantee delivery of all records to the target in the event of restart or cluster failover.

Record Filtering

Capture performs table and row filtering based on configuration criteria, and filters out unnecessary administrative log records when processing transaction logs.

Column Mapping and Transformation

Capture can select specific columns and apply column transformation rules before sending data to the target. Transformations can be specified via built-in functions, user-supplied code or stored procedures.

Changes Written to Queues

After applying record filters, column mapping and transformation functions, Capture outputs data to one or more platform queues. GoldenGate platform queues are a set of flat files maintained by the Manage component that enable decoupling of Capture and Deliver activities. Capture can create unique, as well as overlapping, sets of data into each queue.

Queued records are typically stored in a proprietary GoldenGate format, but can also be stored in XML and other formats compatible with different applications. Queued records are usually written to binary flat files on the target system, but may also be written to the source system, or to alternative queue types and application interfaces.

DELIVER

GoldenGate's Deliver module enables a wide range of data synchronization and backup processes across nearly all major databases. Its powerful capabilities enable customers to:

- Read queues for the latest database changes
- Execute mapping and transformation rules according to custom specifications
- Apply changes to designated tables in target databases

Changes Retrieved from Queues

As soon as data is available, Deliver retrieves inserts, updates, deletes and transaction information from a designated queue created by Capture. Deliver checkpoints its position in the queue, which guarantees posting of all records to the target database.

Table and Column Mapping

Deliver enables administrators to specify which source tables they want to replicate. As with Capture, specific records or columns within the table can be selected via user-defined criteria. For each source table, administrators also specify at least one target table as the destination of synchronization. More than one target table can be specified per source. By default, Deliver populates any target table column with data from a source table column if the two columns share the same name. Explicit mapping rules and transformation rules can also be specified, from simple column assignment rules to more complex transformations. Implicit mapping and explicit rules can be combined.

Changes Applied to Target Database

If selectivity rules are passed, Deliver posts each database change to one or more target tables. Changes are applied in the same order as they were committed in the source database to ensure data and referential integrity. In addition, changes are applied within the same transaction context as on the source system to ensure read consistency on the target. Deliver utilizes native database calls, statement caches and local database access to execute updates efficiently.

MANAGE

The Manage module enables users to exercise complete administrative control over all GoldenGate components and synchronization functions across the network. Employing a familiar, browser-based Activity Console, the Manage module is an effective tool for:

- Proactively monitoring performance
- Initiating component processes and executing processing requests
- Preventing disk-full conditions
- Viewing and processing events
- Identifying problems before they can affect performance

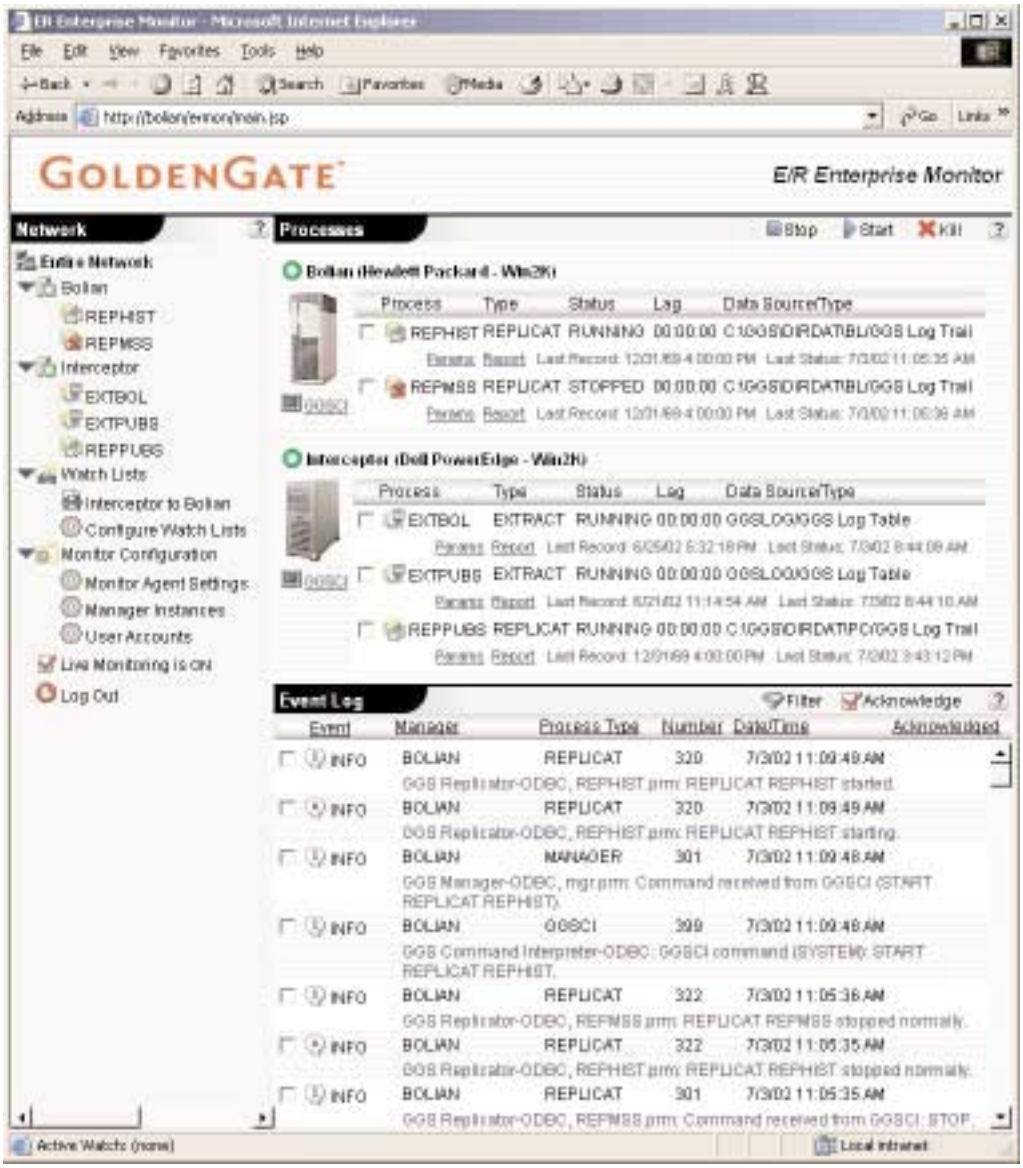


Figure 4 - Manage Main Console

Process Management

GoldenGate's Manage component executes processing requests both on demand and automatically. Manage starts and restarts GoldenGate components, and monitors synchronization latency in order to provide proactive alerting to potential problems before they become critical.

Automatic Queue Management

Manage automatically truncates queue data when no longer needed, ensuring against inadvertent disk-full conditions and providing an alternative to error-prone manual procedures. In addition, Manage provides parameter-driven queue retention functions, which enable flexible data management and recovery options.

Optional Interface

As an alternative to its Web browser interface, GoldenGate also provides a standard command-line interface. In addition, Manage enables monitoring and control of modules and events via batch and shell scripts.

Event Management Options

Processing events can be displayed by GoldenGate interfaces and through standard event management tools such as Tivoli and HP OpenView.

ADDITIONAL CAPABILITIES

Data Synchronization and Application Integration Features

Table, Row and Column Selectivity

GoldenGate provides a wide range of options to pinpoint the data that is desired for synchronization. Specific tables, rows and columns within tables can be selected for synchronization.

Column Mapping and Transformation

Tables with dissimilar structures can be replicated to one another. GoldenGate provides implicit, automatic mapping of columns with common names in source and target tables. In addition, GoldenGate enables explicit mapping and transformation capabilities between columns. Using built-in date, math, string and utility functions, transformations can be specified and columns can be combined and divided.

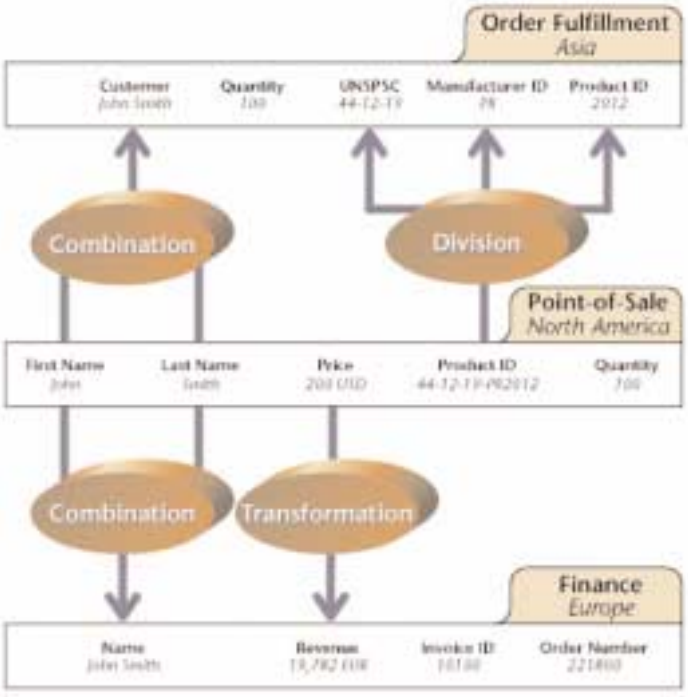


Figure 5 - Mapping and Transformation

Archival and Audit Functions

GoldenGate automatically creates an archive of purged information from the source database by transforming delete records into inserts. For auditing and related purposes, GoldenGate can also maintain a separate history table to track each update to individual records as records change.

Application Interfaces

GoldenGate provides data as it changes to user-written programs, database stored procedures, MQ Series queues, XML, ASCII and EBCDIC formatted files, IBM xSAM files and other destinations. These alternatives provide a variety of options for real-time data synchronization and application integration. Custom processing can also be utilized to perform complex mapping and selectivity functions.

Scheduling Options

GoldenGate performs synchronization in coordination with other application activities. GoldenGate can be driven via shell scripts and similar mechanisms, and can be run in a periodic batch cycle as well as in perpetual mode.

Flexible Synchronization Topologies

To support underlying application requirements and architectures, GoldenGate supports a variety of synchronization topologies. These include one-to-one, one-to-many, many-to-one and cascade synchronization scenarios. Peer-to-peer synchronization is also supported via automated synchronization loop detection. In addition, specific sets of database changes can be staged on the source system, on the target system, or both. Each set of staged data can contain unique or overlapping sets of data.

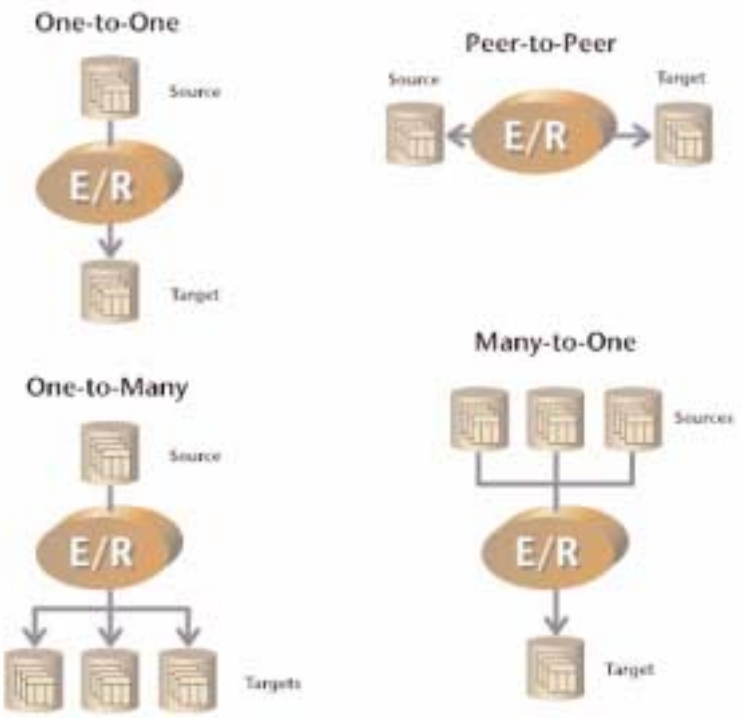
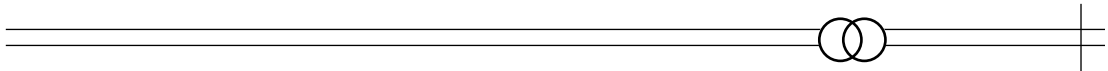


Figure 6 - Flexible Synchronization Topologies



Selective Backout
GoldenGate provides the capability to reverse selected transactions, and operations within transactions. Reversal criteria can include specific tables, column values, operation types (such as deletes) and time ranges.

Conflict Detection and Resolution

Ensuring Data Accuracy
GoldenGate takes the approach that the fundamental issue underlying conflict detection and resolution is data accuracy. Reflecting this thinking, our solutions provide a number of capabilities that ensure data accuracy in complex environments and thereby avoid conflict.

Resiliency
- Synchronizes data automatically
- Persistent processing when network, system errors occur
- Cluster-compatible for transparent failover
- Immediate backup of data and change queues

Data Integrity
- Ensures that data is processed exactly once, in the proper order
- Maintains transaction properties and referential integrity (explicitly and implicitly)
- Unresolved conflicts result in transaction rollback

Self-managing and Proactive
- Pro-active threshold management for early warnings that enable the resolution of problems before they happen
- Self-maintaining queues preserve resources
- Minimized queue space requirements to tolerate extended outages

Performance
- Minimal processing requirements ensure low latency, minimize queueing and maximize disk cache utilization while maintaining prior application service levels
- Supports low latency even at high transaction volumes, dramatically reducing the chance of "simultaneous" updates on the same row

Diversity
- Basic and advanced conflict detection enables discernment, for both actual and nominal conflicts between similar and dissimilar databases
- Basic and advanced resolution mechanisms accommodate application-specific requirements, mul tiple data sources, similar and dissimilar databases

Audit and Recovery

- After-the-fact audit and resynchronization
- Granular, custom recovery capabilities
- Transaction history
- Roll-forward target databases using source changes

Flexible Resolution Options

When encountering conflict during synchronization, GoldenGate provides a number of resolution options, including the ability to bypass or retry an operation, insert missing rows, identify conflicts in "dissimilar" data, combine column values from different sources, log an operation into an exceptions table, report conflicts, terminate processing, and many other options.

Transaction History Databases

GoldenGate provides the ability to create complete and flexible databases for auditing synchronization activity, making it possible to track columns that changed in each database operation, which data source changed the data, where the data was routed, and the linkage between the different changes (rows) associated in a particular transaction.

Reliability, Integrity and Security

Data Integrity, Transaction Consistency

GoldenGate applies replicated operations in the same order that they are committed on the source system. In addition, replicated operations are applied in the same transaction context on the target as on the source, ensuring read-consistency is preserved on secondary systems for accurate query and transaction processing.

Online Synchronization

Source and target tables can be synchronized without taking the source database off-line, enabling a complete refresh of the target to occur without interrupting application services.

Change Data Management

GoldenGate stages changed data into intermediate queues prior to posting on the target system. The staging queues allow decoupling of source and target system processing and enable synchronization services to accommodate occasional system, database and application maintenance activities with full continuity. The number of GoldenGate queues is minimized, and queue data is deleted automatically, in order to eliminate associated administrative tasks and ensure data integrity.

Robust Data Security

GoldenGate provides 128-bit encryption utilizing the Blowfish algorithm from Counterpane Internet Security to ensure secure, confidential data transmissions. Blowfish is a symmetric block cipher that can be used as a drop-in replacement for DES or IDEA. GoldenGate's implementation of Blowfish can take a variable-length key, from 32 bits to 128 bits, making it ideal for both domestic and exportable use.

Guaranteed Data Delivery and Synchronization

GoldenGate components feature reliable and efficient checkpointing, which ensures guaranteed delivery and synchronization of data. GoldenGate automatically withstands network and system outages.

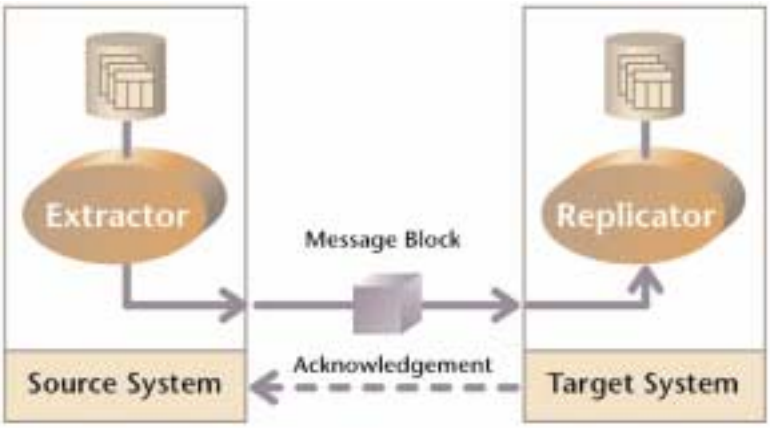


Figure 7 - Guaranteed Delivery Queue

GoldenGate Making a Difference Worldwide

GoldenGate delivers Global Data Synchronization - the immediate capture, transformation and delivery of large volumes of data across your entire organization and beyond. For the real-time enterprise, GoldenGate provides the essential technology and expertise to implement a wide range of business and IT solutions, from disaster recovery and business continuity to application integration, system migration and active data warehousing. More than 200 organizations rely on GoldenGate to run their businesses in banking: Bank of America, Bank One - financial services: VISA, Merrill Lynch - healthcare: Mayo Clinic, Legacy Health System - and technology/communications: Comcast Communications, AOL/Time Warner. Where instant access and synchronization of up-to-the-second business information is essential, GoldenGate is the platform. For more information, please visit the company's web site at **www.goldengate.com**