Informatica Data Replication
Moving and Synchronizing Real-Time Data
in a Heterogeneous Environment
This document contains Confidential, Proprietary and Trade Secret Information ("Confidential Information") of Informatica Corporation and may not be copied, distributed, duplicated, or otherwise reproduced in any manner without the prior written consent of Informatica.

While every attempt has been made to ensure that the information in this document is accurate and complete, some typographical errors or technical inaccuracies may exist. Informatica does not accept responsibility for any kind of loss resulting from the use of information contained in this document. The information contained in this document is subject to change without notice.

The incorporation of the product attributes discussed in these materials into any release or upgrade of any Informatica software product—as well as the timing of any such release or upgrade—is at the sole discretion of Informatica.

Protected by one or more of the following U.S. Patents: 6,032,158; 5,794,246; 6,014,670; 6,339,775; 6,044,374; 6,208,990; 6,208,990; 6,850,947; 6,895,471; or by the following pending U.S. Patents: 09/644,280; 10/966,046; 10/727,700.

This edition published May 2013
# Table of Contents

**Executive Summary** .......................................................... 2

Data Replication Challenges .................................................. 3

**Key Data Replication Criteria** ........................................... 4

Informatica Data Replication .................................................. 5

Data Replication Usage Scenarios ............................................ 6

Real-Time Data Warehousing .................................................. 6

Offloading to a Reporting, Query, and/or Analytics Server .......... 6

Real-Time Data Integration ................................................... 7

Continuous Availability ....................................................... 7

Real-Time Auditing ............................................................. 8

Informatica Data Replication Architecture ................................ 8

Console ............................................................................. 8

Server Manager ................................................................. 9

Initialsync ................................................................. 11

Extractor ................................................................. 11

Applier ................................................................. 11

Recovery and Rollback via Checkpoint Processing ................. 12

Supported Topologies of Informatica Data Replication ............ 13

Conclusion .......................................................... 17
Executive Summary

IT organizations in all vertical industries are under pressure to handle growing data volumes and increasing demands for better operational insight and intelligence. To improve operational decision-making and reduce downtime related to migrations and upgrades, IT organizations must replicate data continually and efficiently.

But there are inherent challenges. Data replication requires not only speed to replicate data as quickly as possible but also the ability to scale to handle enormous volumes of data.

To be truly effective for real-time data warehousing and large-scale, high-speed operational synchronization, data replication software must have certain capabilities. It must provide native support for a broad range of heterogeneous source and target data stores and systems and for multithreading and parallelism. It must also guarantee transaction integrity for consistent data across all systems.

But even beyond speed and scale, data replication software must also provide other important capabilities:

- Robust support for multiple data distribution topologies
- Centralized design, deployment, and monitoring of configuration topologies through a robust graphical user interface (GUI)
- Change data-capture without intrusion on source systems, enabling replication of changes only without duplication of the entire database
- Persistent intermediate storage of changes to make sure that data is received and committed to the target database even in the face of interruptions or outages

Informatica® Data Replication delivers all these capabilities. This high-performance enterprise software moves and integrates data in real time. It extracts data from wide variety of sources in a highly efficient way, and then quickly loads data into target destinations. It easily scales to handle large data volumes.

With Informatica Data Replication, your IT organization can easily and efficiently move and synchronize data across the enterprise to improve data currency, quality, availability, and continuity. And because this real-time capability is part of the Informatica Platform, you can dramatically streamline your entire data integration process across both conventional, batch-oriented data and real-time, current data.

This white paper discusses how new real-time data management challenges have placed new and higher demands on data replication. It outlines the criteria your IT organization can use for selecting data replication technology. It then describes how Informatica Data Replication meets, and in some cases exceeds, these criteria. It explains how Informatica Data Replication is used in common data replication scenarios. This paper also examines the architecture and supporting topologies of Informatica Data Replication, ultimately validating that the software is the strategic choice for real-time enterprise data integration.
Data Replication Challenges

Data replication has been essential to IT organizations for decades. It first emerged in the early 1990s to address the limitations of database technologies. It allowed transactional operations across distributed copies of the database. As data replication was adopted as a distributed data movement solution, it became apparent that enterprise-class replication products could address other critical data movement challenges, such as:

- Continuous and high availability during maintenance operations
- Fault resilience during unplanned outages
- Load balancing through offloading up-to-date copies of data for business intelligence, reporting, and ad hoc query activity from production systems

For more than two decades, data replication has remained an effective, multipurpose data movement solution. As a result, many analyst firms and industry observers rank data replication as the number two data integration technology in use today, after extract, transform, load (ETL).

New business demands for real-time data have placed new demands on IT organizations and, as a result, on data replication technology:

- Exponential growth in data volumes in both operational and analytical/warehousing systems
- Mandatory operational business intelligence, requiring real-time data delivery to all warehousing and analytical systems as well as operational ones
- Increasing heterogeneity of both operational and warehousing systems, including mainstream use of new data types and data stores, from analytical data warehousing appliances to new data storage systems such as Hadoop, in addition to traditional relational, transactional databases
- Sentiment analysis, pattern detection from sensor-based data, and other new types of analysis that must be incorporated into real-time, actionable decision making in operational systems and functions

Today's data replication technology must offer a wider range of capabilities. It must satisfy both traditional and emerging use cases while maintaining high performance, low latency, data and transaction integrity, and reliability—all with little-to-no impact on source system performance. And it must be able to deal with an ever-expanding range of traditional and emerging source and target systems, including data warehouse appliance solutions such as Oracle Exadata, EMC Greenplum, HP Vertica, IBM Netezza, and Teradata, as well as distributed file systems such as Hadoop. To support these newer systems, replication deployments must provide innovative techniques to optimize transaction loading into each, beyond the traditional structured query language (SQL) apply techniques that currently prevail in replication technologies.
Key Data Replication Criteria

To address the current range of business requirements, data replication software must address the following key criteria:

• High performance and scalability, delivering a very high volume of records from hundreds or thousands of tables simultaneously with little-to-no latency. It should maintain this delivery in near-real-time despite growing volume, diversity, and dispersion of the source and target data systems.

• Support for multiple topologies, including one-to-many, many-to-one, many-to-many, and active-active (bidirectional) configurations and the flexibility to allow multiple delivery options ranging from continuous to delayed, as well as through messaging and traditional networking.

• Reliability and continuous availability of business systems and data, with minimal performance impact or service interruption during planned maintenance, such as migrations and upgrades, and unplanned outages.

• Transaction integrity for maintaining the entire scope of a transaction as its data is moved between source and target systems.

• Security that enables all data to be encrypted and protected during movement between systems.

• Minimal overhead and impact on production systems with nonintrusive data capture from the source systems. Log-based change data capture is most desirable.

• Ease of integration and use through a GUI that handles all aspects of design and deployment, including mapping among different data stores and types. All implementation should be accomplished via the GUI for the easiest and fastest time to value. Replication software should also integrate easily with existing systems, platforms, and databases and not require any change to these systems.

• Centralized management through a comprehensive console for monitoring and managing the entire configuration, regardless of complexity or geographic dispersion.

• Heterogeneity and versatility in covering the range of popular databases, warehouses, appliances, operating systems, and platforms in use and in addressing emerging data stores, such as Hadoop and analytical appliances, as well as traditional relational database management system (RDBMS) stores. Rather than limit users to any single vendor’s database or data warehouse, the software should remain agnostic. It should also provide optimized delivery (apply) capabilities for each type of target system supported and the ability to capture schema, as well as data changes among heterogeneous databases.

• Support for a broad range of use cases in distributed data synchronization, including real-time data integration, real-time operational business intelligence and data warehousing, continuous availability, query and report offloading, transaction audit capture, and zero downtime maintenance and migration support.

• Collaboration and true integration with other data integration technologies, such as transformation, data quality and governance, data virtualization, high-speed messaging, complex event processing, and delivery into the cloud. It should participate in and with these capabilities in an integrated fashion while maintaining consistency with what a replication solution is intended to provide.
Informatica Data Replication

Informatica Data Replication meets all the key selection criteria for an enterprise-class data replication product offering. Its highly flexible technology allows organizations to capture, route, transform, and apply transactional data among a full range of heterogeneous databases in real time with little to no impact on the source systems. It can be applied to both traditional and emerging business requirements and use cases where information is moved among databases, applications, and systems. Informatica Data Replication is a single, manageable, high-performance product that meets a broad range of current and future enterprise data movement needs.

Informatica Data Replication offers many distinct advantages over other replication solutions in the industry, including:

• Fast time to value. The Informatica Data Replication GUI console allows you to design, configure, map, and deploy the entire replication configuration in a straightforward fashion.

• Optimization for real-time data warehousing. Informatica Data Replication’s unique merge apply capability optimizes transactional data delivery into warehousing targets and analytical appliances and applies the data seamlessly, without the need to run generated scripts. Most replication solutions offer only SQL Apply capabilities, whereas Informatica Data Replication offers three different apply techniques: SQL apply, merge apply, and audit apply.

• True heterogeneous support. Informatica Data Replication supports a broad range of source and target systems, including warehousing, analytical appliances, and big data stores such as Hadoop. Additional sources and targets are planned. The product provides optimized delivery to each type of target system and allows schema—data definition language (DDL)—as well as data changes—data manipulation language (DML)—to be replicated among heterogeneous relational databases. As independent data integration vendor with no database management system (DBMS) offerings of its own, Informatica is committed to providing its customers with the full range of data stores in Informatica Data Replication.

• Comprehensive topology support. Informatica Data Replication supports a full range of replication topologies, from simple, one-to-one replication to complex, bidirectional and many-to-many configurations. The software also provides full conflict detection and resolution capabilities to manage update collisions and maintain the integrity and consistency of data across bidirectional implementations.

• Live audit for data governance, compliance, and risk management. Informatica Data Replication’s audit apply is another unique capability for replication solutions. Audit apply captures all transaction activity and formats it for easy search and analysis to improve insight into the transactions being initiated and how they are performing.

• Integration with the Informatica Platform. As part of the broader Informatica Platform, Informatica Data Replication allows IT organizations to invest in a comprehensive set of advanced data integration capabilities that work together synergistically to meet all data quality, currency, movement, and integration needs.

Informatica Data Replication is highly scalable, reliable, and easy to configure. It allows IT organizations to share information across different systems in a heterogeneous environment by replicating data among different hardware platforms and data sources while maintaining its transactional integrity. The software performs low-latency batched or continuous replication of data changes as well as metadata/schema changes from source databases to target databases over a LAN or WAN. Additionally, alternate transport mechanisms can be
utilized to replicate data and/or schema changes by leveraging an external API (which is provided) to publish to either a Java Messaging Service (JMS) messaging queue, or to Informatica’s high-speed Ultra Messaging bus or to any other Enterprise Service Bus (ESB) of choice. And because Informatica Data Replication uses log-based change data capture (CDC) techniques, the impact on the underlying database systems and user applications is minimized.

Informatica Data Replication includes an efficient bulk data load tool that can be used to load data from a source database to a target database before starting data replication. Another option for initial load and target synchronization is Informatica Fast Clone, a high-speed, high-compression database-cloning product. Informatica Fast Clone currently supports Oracle as a source and all of the Informatica Data Replication target databases and appliances. Informatica Fast Clone also provides a direct data streaming capability into Teradata and Greenplum.

Data Replication Usage Scenarios

Informatica Data Replication can be used in virtually any scenario where you need to replicate data among homogeneous or heterogeneous databases and systems. The most common use cases for data replication are:

- Real-time data warehousing for operational business intelligence and real-time analytics
- Offloading reporting, query, and/or analytics activity to a separate server
- Real-time data integration for load balancing, consolidation, and synchronization
- Continuous availability, including near-zero-downtime migration, upgrade, and maintenance support
- Real-time auditing of transaction activity for compliance, risk management, and data governance

Real-Time Data Warehousing

Data warehousing, especially as a centralized concept, is undergoing radical change. Loads are increasing on every level: volumes of data, users, and computations to support analytics and data discovery all continue to grow exponentially. Many IT organizations are moving to distribute their warehousing and analytics platforms, just as they have distributed their operational systems. Analytics and operational business intelligence are now as mission-critical as many operational business applications and require data to be available and up-to-date 24x7. Batch windows are shrinking and for many organizations are now virtually nonexistent. Thus, ETL and extraction, load, and transfer (ELT) alone cannot meet the need. Real-time update in place is now essential, regardless of whether the data is being warehoused or sandboxed in a traditional DBMS, appliances, or Hadoop. Data replication allows the data warehouse to take on any form or distribution but still reflect the latest state of the information for true operational business intelligence and actionable decision making.

Offloading to a Reporting, Query, and/or Analytics Server

Although DBMSs have been optimized for decades to support mixed workload environments, there is no question that compute- and search-intensive activities such as ad hoc querying, analytics and data discovery, and reporting impact the overall performance and throughput of transactional applications running on the same system. Offloading the data for these decision support activities onto separate servers has been a common solution to this dilemma for many years, and it remains a cost-effective solution because it also avoids the need to continuously upgrade and add capacity to operational systems. Today, organizations are recognizing the benefits of offloading reporting and analytics from data warehousing systems as well. With replication, the data in these lower-cost systems is kept synchronized in real time, enabling not just cost reduction but also greater ROI via improved quality, currency, and relevance of analytics and decisions.
Real-Time Data Integration
This is the classic use case set for which data replication emerged as a solution in the early 1990s, and it remains prevalent today. It includes a number of distinct categories:

Data Integration
Data from multiple source systems is integrated, and the result is replicated to one or more destination targets. This technique is often used for creating reference data to be used at remote sites and should not be confused with ETL-style data integration, where data is significantly transformed or aggregated for analytic use.

Data Consolidation
Data from multiple source systems is integrated and consolidated onto a single target system. Data warehouses and operational data stores (ODSs) for business intelligence (BI) systems are one common consolidation use, but there are operational system examples as well, such as providing corporate headquarters with a consolidated view of inventory levels across various distribution sites.

Data Distribution and Synchronization
Data is distributed among multiple systems and kept synchronized with any transaction activity occurring against that data initiated from any site. The distributed data, or the update activity, is usually partitioned in some fashion to prevent transaction collisions and to avoid the need to resolve conflicts. In rare cases, all sites operate as multiple masters or peers and can update any portion of the database. In the latter case, it is important that the replication solution selected provides conflict-handling mechanisms to manage transaction collisions and maintain integrity.

Continuous Availability
The cost of downtime and system outages has never been higher than now, when businesses must operate around the clock and provide up-to-date information 24x7 to remain competitive. The costs can range from lost revenue and business opportunity to damaged relationships with customers and partners and even legal exposure, such as breach of service. Businesses simply cannot afford downtime of any type.

Downtime and unavailability can take several forms:

- Planned Downtime. Systems and databases must be maintained throughout their lifecycle, whether maintenance is in the form of test systems moving to production, applying fixes/patches, upgrading versions, or migrating across platforms. Application and data availability must be maintained with the nearest-to-zero downtime impact of any such operations.
- Unplanned Downtime. Unplanned downtime can range from common, short-term system or database unavailability due to network outages, system failures, or software bugs to rarer but more significant disasters that can impact the business for a much longer time.

Data replication reduces the risk from both planned and unplanned downtime by providing one or more additional, synchronized copies of a full database or subset. Applications and data access can continue on one system while the other is being maintained, upgraded, or migrated. All transaction activity is captured and resynchronized when off-line systems are brought back online. Applications and activity therefore continue without interruption and without loss of recent data updates.

Creating one or more full replicated copies of the database in a remote location is a common way of enabling disaster recovery. Using data replication, all transaction activity is replicated to the off-site copy, minimizing the impact of any major disaster by providing a standby database to which applications and activity can be redirected.
Real-Time Auditing

All transaction activity on a database or subset is captured from the logs and persisted independently of the database in question. This is done in a format that allows querying and reporting to determine whether any activity violates compliance directives or introduces risk to the organization.

Informatica Data Replication Architecture

Informatica Data Replication has several features that run as separate, decoupled executable functions (see Figure 1). These features can be flexibly configured across the enterprise to provide maximum performance and availability:

- **Console**
  - Extractor
  - Server manager
  - Initialsync
  - Applier
  - Recovery and rollback via checkpoint processing

![Figure 1. Informatica Data Replication is heterogeneous and versatile, consisting of several functional components that run as separate, flexibly configurable executables.](image)

**Console**

The Informatica Data Replication console is a GUI for performing functions on replication jobs, including configuration, administration, deployment, monitoring, scheduling, and management. The console can be run on its own on any computer system that is connected to a source and target system. It has been designed for multi-user access and provides enhanced access security features, which prevent users from accessing or editing objects outside of their own administered privileges.

**Flexible Table and Column Selection, Mapping, and Transformation**

The console can be used to define source and target database tables as well as table and column mappings. It allows full flexibility in selecting the data updates to be replicated from the source to the target and in determining the format of the target schema (see Figure 2). And with its schema generation option, it enables compatible target table definitions to be made based on selected source table definitions. The console leverages Java database connectivity (JDBC) to extract the metadata required for mapping and execution of DDL. Simple transformations can be made in Informatica Data Replication via scripted functions either in SQL or TCL. The software also allows more extensive or complex transformations when used with Informatica PowerCenter®.
Powerful, Yet Intuitive and Simple Operation
The console’s GUI is very simple and straightforward. All table and column mapping is executed via the GUI, including adding any filtering conditions based on column values. Full monitoring is built in, including support for simple network management protocol (SNMP) with real-time alerts, transaction history, heartbeats, and historical statistics for extractor, applier, and send file processes. Full support for import and export is also included. Figure 3 shows some of the statistics and status information displayed by the console.

ICC-Class Centralized Deployment
The console allows database administrators (DBAs) to deploy configurations of any topology across multiple server managers from a single point, making the task of deploying configurations much easier and more manageable. It also streamlines the process of promoting a configuration from development to test to production status.

Server Manager
The server manager of Informatica Data Replication acts as the secure, central gateway for configuration, management, and monitoring of all replication jobs and enables data movement from a source database to a target database over a LAN or WAN. Its flexible and advanced configuration management and monitoring capabilities enable streaming of data to multiple target systems while maintaining data consistency.
The console of Informatica Data Replication connects to the server manager to perform the following tasks:

- Create, import, edit, or export configuration files
- Define and manage connection definitions
- Schedule extractor, applier, initialsync, and send file tasks
- Monitor replication tasks
- Configure user notifications by email

Data replication includes the flexibility to schedule replication jobs to meet specific performance and latency requirements:

**Ultra-Low-Latency Continuous Replication**

Data replication can perform continuous replication of data in near-real-time for all supported sources, targets, and apply modes. Continuous replication jobs capture and transmit data to intermediate files in configurable short-latency microcycles, usually consisting of milliseconds. The console is used to schedule continuous replication jobs to run with continuous periodicity and in a time period of up to years in the future.

Figure 3. The statistics-latency screen of the console displays real-time and average latency for the extractor, applier, and send file processes.
Scheduled (Deferred) Replication
Informatica Data Replication can also run replication tasks within a specified time and at a specified frequency. The time and frequency are defined when a replication task is scheduled using the server manager’s built-in scheduler in the console. The frequency interval can be seconds, minutes, hours, or days. The time can be defined by specific start and end dates and times.

In a distributed Informatica Data Replication topology, the server manager can be run on multiple machines. Typically, a server manager instance is run on every machine where an initialsync, extractor, or applier component runs. One server manager acts as the main server and the others as subservers. The main server can run on any of the machines in the replication topology.

Initialsync
The initialsync feature of Informatica Data Replication materializes target tables from source tables. It is used to perform an initial load of data prior to starting replication. The initialsync process can also be parallelized to ensure optimal performance during initial capture of the source database contents. Depending on requirements, the capture can include either a subset or the entire contents of the source database.

As noted previously, Informatica Fast Clone can also be used to perform the initial load/materialization.

Extractor
The extractor feature is the change data capture component of Informatica Data Replication. It reads the transaction logs of the source database system and extracts data changes, such as inserts, updates, and deletes, for the selected source tables defined in the configuration file. The extractor feature can also detect and capture the schema/DDL changes and then propagate those identified changes into the target system.

High Volume with Minimal Impact
The extractor captures all relevant database changes directly from the database logs, which are maintained and managed by the DBMS itself. The extractor does not create objects in the source database that generate ongoing I/O. As a result, the source DBMS is minimally impacted by the extraction process. The extractor is tuned for high performance with parallel execution capabilities and can handle even the most demanding volume requirements. It can also parse logs produced on other architectures and can extract to text files that can be processed by Informatica PowerCenter, other ETL tools, or other third-party applications.

Intermediate Files Further Minimize Overhead While Providing Full Data Security
The extractor stages all data changes in compressed on-disk intermediate files, again minimizing impact on DBMS resources while providing complete decoupling of source and target systems to support full heterogeneity. All data in the intermediate files is encrypted using open secure sockets layer (SSL) encryption, helping to ensure that the data is kept secure and confidential as it is routed from source to target.

The extractor is usually run on the source system, but it can be run on another system if it can access the transaction logs. In addition, the logs can be shipped or transferred to the system where the extractor is running.

Applier
The applier is the delivery component of Informatica Data Replication. The applier reads only committed changes from the staged intermediate files and applies these changes to the target. After the data is committed, it is purged from the intermediate files.
Transaction and Data Integrity and Consistency with High Performance for High Volume

Only committed transactions are applied to the target system and in the same order in which they were applied to the source database. This helps to ensure that full transaction consistency and data integrity are maintained. The applier combines all changes from committed transactions since the last apply operation and applies these changes to the target in a single transaction, increasing performance and efficiency by minimizing the number of transactions that are processed on the target. The applier is fully multithreaded and uses automatic, intelligent parallelism that maintains the order during the parallel operation. The parallelism is determined via advanced load balancing algorithms that optimize the Apply task.

Latency can also be customized to meet performance and information delivery requirements by selecting either continuous, ultra-low-latency replication or scheduled replication as discussed in the section that described the server manager.

The applier is usually run on the target system but can be run on another system that can access the target database by using the target’s connectivity client or open data base connectivity (ODBC).

Three Different Apply Modes Optimize Delivery Performance

The applier provides the following apply modes: SQL apply, audit apply, and merge apply.

1. SQL Apply

SQL apply executes SQL statements on the target that are equivalent to the SQL change operations on the source. This is the most prevalent mode in data replication solutions today. As noted earlier, Informatica Data Replication optimizes this apply by combining all changes since the last apply operation into a single transaction, a higher-performance option than simply a direct replay of all insert, update, and delete statements.

2. Merge Apply

Merge apply is an optimized apply capability for replication to analytical/data warehousing appliances including Greenplum, Netezza, Teradata, and Vertica. For each source table, a staging table exists on the target. Changes are bulk loaded into the staging table, and the data is merged as a mini-batch from the staging table into the actual destination table.

The SQL merge statements are optimized for the types of SQL changes in the apply batch. For example, if the apply batch contains only inserts for a table, the merge operation uses a simple INSERT INTO SELECT statement. If the apply batch contains updates and deletes for a table, two joins are performed on the target table: one join for the updates and one join for the deletes.

3. Audit Apply

Audit apply is used to audit changes for regulatory or oversight purposes. With this apply mode, each source table has a corresponding audit log table on the target. The audit log table records a row for every SQL update, insert, or delete on a source table. Each row contains a before- and an after-image of the data for each column in the table and for the columns that Informatica Data Replication adds to store metadata about the operation. The metadata includes the SQL change type, timestamp, SCN, transaction ID, commit SCN, and commit time.

Recovery and Rollback via Checkpoint Processing

Informatica Data Replication includes powerful capabilities using persisted checkpoints to recover from exceptions that disrupt replication. The software records checkpoint information separately for the extractor and applier, storing each checkpoint in the same transaction scope as the data it places in each target’s
recovery table. It can then use the checkpoint information to provide data consistency across all of the replication stages and in the event of failures. Checkpoint information is persisted externally as well as in the target recovery tables. As a result, checkpoint information will not be subject to database failures at the source or target.

The extractor, server manager, and applier of Informatica Data Replication use checkpoint and recovery information to resume processing and prevent inconsistency after failures. If an Apply operation fails, the entire batch of changes is rolled back on the target database. If the applier is restarted or the database is shut down and restarted, the applier will accumulate the changes and restart from the last checkpoint.

Supported Topologies of Informatica Data Replication

Informatica Data Replication’s modular architecture supports the complete range of data distribution topologies, from the simplest to the most complex. Its components flexibly adapt to any network configuration. It can be deployed on a single system for local replication or on different systems for distributed replication across an organization’s full computing environment.

The most common topologies are:

• One source to one target
• One source to multiple targets
• Multiple sources to one target
• Multiple sources to multiple targets

Figure 4. Informatica Data Replication can be configured for one-to-one replication.
One Source to One Target
This topology is also known as 1:1, point-to-point, one-way, or unidirectional (see Figure 4). Common use cases for this topology include:
• Offloading query, reporting, and/or analytic activity to a separate server
• Providing continuous availability and zero downtime during migrations, upgrades, and maintenance
• Supplying a warm standby database copy for fault resilience

One Source to Multiple Targets
This topology is also known as 1:M or broadcast (see Figure 5). Common use cases for this topology include:
• Data distribution to support data access that is geographically, organizationally, or operationally dispersed
• Another level of fault resilience via copies in multiple locations
Multiple Sources to One Target
This topology is also known as M:1 or consolidation (see Figure 6). Common use cases for this topology include:

- Real-time updates to a data warehouse or operational data store
- Geographically, organizationally, or operationally distributed report-back to a central consolidated ODS

**Figure 6. Informatica Data Replication can be configured for many-to-one replication.**
Multiple Source to Multiple Targets
This topology is also known as M:M, peer-to-peer, active-active, or bidirectional (see Figure 7). Common use cases for this topology include:

- Load balancing via partitioning the data and/or the transaction activity
- Multimaster or bidirectional where updates/transactions can occur on the same data. This case requires conflict handling.

![Data Replication Console](image)

The console runs on a standalone client.

Informatica Data Replication
Tasks:
- Server Manager
- Extractor
- Applier Task
- Database client software or ODBC

Figure 7. Informatica Data Replication can be configured for bidirectional, many-to-many replication, with full conflict handling.

Cascade Replication
Cascade replication is a hybrid topology that replicates changed data across a series of databases. In this topology, Informatica Data Replication uses one or more databases in the middle of a series as source and target simultaneously. Figure 8 shows a typical cascade replication topology that includes three databases.

In this topology, changes may occur on each database, but Informatica Data Replication replicates only the changes that initially occur on each database, skipping those replicated from another database. For example, for Database C, Informatica Data Replication skips transactions replicated to Database B from Database A, applying only the changes originated from transactions on Database B.
Conclusion

Your company faces increasing competitive pressure, compounding expectations on your IT organization. Explosive data growth has boosted demand for real-time analytics, operational business intelligence, and continuous availability. Your team needs to integrate disparate systems in real time, replicate reporting and operational information, and augment data warehouse feeds with new and updated data with the lowest possible latency.

Your IT organization needs a new kind data replication technology—one that provides a wide range of capabilities to help your team handle not only traditional data replication use cases but emerging ones. This new kind of data replication technology must offer high performance, low latency, data and transaction integrity, and reliability—all with little to no impact on source system performance. And it must scale to handle an ever-expanding range of source and target systems.

Informatica Data Replication meets all these needs. With it, database administrators as well as business intelligence and data warehouse IT professionals can accelerate project delivery, improve the performance, and lower costs. Informatica Data Replication empowers your IT organization to:

- Enhance agility with real-time data warehousing for operational business intelligence and real-time analytics
- Reduce impact on operational systems by offloading reporting, query, or analytics activity to a separate server
- Boost efficiency by facilitating real-time data integration, consolidation, and synchronization to ensure that all data in distributed systems remains synchronized and accurate
- Increase productivity by providing business users with a complete 360-degree view of business entities
- Save costs and perpetuate data availability by minimizing downtime from planned and unplanned outages

Informatica Data Replication’s robust features and capabilities, architecture, and supporting topologies make it the strategic choice for real-time enterprise data integration.