

# Data Replication for Operational Reporting

## Executive Summary

Every business intelligence specialist has noticed it: the world of BI is changing. On one hand, new technologies, such as data warehouse appliances, self-service BI tools, and NoSQL database servers, are creating new opportunities to transform data into information in ways that were unthinkable just a few years ago. On the other hand, the speed of business has accelerated and with it the need to generate new forms of reporting and analytics to drive informed decision making. Operational reporting has emerged as one of the main alternatives available to IT to support the real-time decision making needs of the business.

In a nutshell, operational reporting supports decision-making processes with operational data (sometimes referred to as real-time data or zero-latency data). This operational data is useful to an increasing number of user groups, including, but not limited to, operational management, the workforce, and external stakeholders, such as customers and suppliers. Implementing operational BI promises to help organizations improve the efficiency of individual business processes and the organization as a whole, reduce operating costs, and increase customer satisfaction.

In many ways, operational reporting and analytics is no different from other forms of BI. End users want to view data as 3D graphs and dashboards, to drill down and roll up the data, and to interact with it in multiple ways. However, what sets operational reporting apart is its reliance on operational data. For the purposes of operational reporting, data is useless when it's a few hours old, never mind a few days. It needs to be as current, as close to real time, as possible.

How can businesses get access to current/fresh operational data in their reporting environment? *Data replication* is the answer. It is a key technology that is being increasingly utilized to provide reporting environments with access to current operational data.

# Operational Business Intelligence: The Business Value of Data Replication

Most current business intelligence architectures offer users access to data that is one day, one week, or maybe even one month old. For a long time, this was adequate for most users. Today, though, users are demanding access to data that is (almost) 100% up-to-date. In other words, they want to work with operational data.

There are many examples of environments that need operational reporting and analytics. A few are highlighted below:

- **Supply Chain Management.** A retail company may want to know whether a truck, already on the road to deliver goods to a specific store, should be redirected to another store that suddenly has a more urgent need for those products. Yesterday's data is too old to perform this analysis.
- **Credit Card Fraud Detection.** A classic form of credit card fraud is when stolen card data is used to purchase products. Each new purchase must be analyzed to see if it fits the buying pattern of the card owner and whether the purchase makes sense. One of the checks could be whether two purchases in different cities occurred within a short time of each other. For example, if a new purchase is made in Boston and the previous one was in San Francisco a few seconds earlier, it's very likely this is a case of fraud. This form of analysis only makes sense on operational data.

As described above, having access to current operational data is paramount in operational reporting. This access to current/fresh data can be enabled by data replication. Data replication facilitates improved decision making by providing access to zero-latency operational data within the BI environment. With access to current operational data in BI environments, your organization will be equipped to operate at the speed of today's business and will no longer be constrained by stale/old data.

## Enabling Operational Reporting with Informatica Data Replication

Informatica® Data Replication is a heterogeneous, real-time transaction replication technology that is highly scalable, reliable, easy to configure, and nondisruptive to the performance of operational systems. It enables operational reporting by providing access to the most current/fresh data in the reporting environment regardless of the complexity and diversity of the IT landscape. As a result, business users can operate with agility and without being constrained by stale information.

Informatica Data Replication offers several key features which differentiate it from other replication technologies:

**Intuitive, Easy-to-Use GUI.** An intuitive GUI supports all aspects of design and deployment, including mapping among different data stores and data types. This GUI makes Informatica Data Replication easy to implement and administer. IT organizations can construct and deploy distributed topologies, as well as monitor the performance and other characteristics of the replication process—all from a centralized location. GUI-driven mapping, configuration, deployment, scheduling, and management capabilities eliminate the need for scripting. You can design once and deploy anywhere across a distributed landscape.

**Unobtrusive Transactional Capture.** Log-based change data capture (CDC) techniques detect changed transactional data and deliver it to target systems with low latency while maintaining the transactional integrity of the data. Because it uses log-based CDC technology, Informatica Data Replication extracts data with minimal impact on the performance of the operational source systems. It accesses database logs instead of using performance-impacting database resources such as triggers or direct SQL queries. Transactional changes are captured in real time, and the changes are applied to other databases, data warehouses, analytical appliances, or even portable format text files, using flexible delivery scheduling.

**Heterogeneous Support and Database Independence.** Informatica Data Replication is a database-agnostic technology. It works seamlessly with an ever-expanding range of traditional and emerging source and target systems, seamlessly delivering data to databases, warehouses, analytical appliances, and big data stores. As more organizations adopt analytical appliances for their reporting needs, Informatica Data Replication provides optimized support for high-performance, transactional, direct-data delivery into data warehouses and appliances such as Teradata, Netezza, Exadata, Vertica, and Greenplum. It can even facilitate loading big data stores by extracting the changed transactional data into a flat file that can then be loaded into Hadoop. It can also load into messaging servers such as JMS. Informatica Data Replication uses an optimized method for quickly loading the data into warehouses and analytical appliances. It can apply light transformations on data in motion by leveraging SQL or TCL scripting before delivering the data to the target system. It supports a full range of replication topologies (e.g., one-to-one, one-to-many, many-to-one, many-to-many, and bidirectional configurations). It promotes automatic cross-platform mapping and facilitates both schema (i.e., DDL) and data replication across different environments.

**Enterprise-Class Scalability and Reliability.** Informatica Data Replication's multithreaded architecture, backed with automatic intelligent parallelism, makes it easy to scale across large volumes. This architecture enables transactional data changes to be delivered while maintaining the consistency and integrity of the data. Reliability of the data delivery process is guaranteed by a unique check-pointing capability. To keep exceptions from disrupting or preventing data delivery into the target system, this capability supports recoverability at the granular, component level. Informatica Data Replication ensures high performance across all transaction volumes. It automatically detects and resolves any transaction collisions by using predetermined rules to allow continuous operations without compromising data integrity. And it supports seamless recoverability from exceptions that may occur during the data replication task.

**End-to-End Integration.** As a component of the Informatica Platform, Informatica Data Replication provides a complete suite of sophisticated data management capabilities. The software works in concert with the core platform to meet all the data integration, data movement, data quality, and data currency needs of an organization.

## Data Replication in Action

This section contains two short case studies of organizations using and benefitting from data replication. Both use Informatica Data Replication to support operational reporting.

**Pharmacy Healthcare Provider** – This large U.S. pharmacy health care provider captures data related to more than 5 million claims each day. Even though its service level agreements called for data latency of less than 4 hours, its infrastructure needed as much as 24 hours to process the operational data necessary for business performance decisions, patient claims, and pharmacy management.

The organization replaced its Oracle data warehouse with a Teradata solution and deployed Informatica Data Replication to replicate data from its Oracle production environment to the Teradata data warehouse. Today, Informatica Data Replication transports approximately 30 tables, each with 30 to 350 columns and many exceeding 100 million rows—and extraction time from the Oracle database is up to 10 times faster than before. As a result, the provider generates reports in faster time for accelerated decision making and faster response to changing market conditions. This enables it to provide better support for pharmacy locations, healthcare payers, and patients alike.

**Telecommunications Company** – A major Asia-Pacific telco replicates its call detail records from its multiterabyte Oracle-based operational data store to a data warehouse developed with the HP Vertica Analytics Platform in 20-minute billing cycles. The system can now alert customers when they are about to exceed the cell phone minutes in their plan, thereby preventing additional expense and avoiding bill shock. Customers can also access their records via a portal tied to the Vertica data warehouse. By giving customers better insight into and control over their voice and data usage, the telco has increased customer satisfaction—not only improving retention rates but also giving customers the incentive to upgrade to plans that let them avoid overages—while generating greater profit.

## Conclusion

Working with data in live environments risks both the data and the applications containing it. Data replication enables operational reporting by extracting zero-latency operational data from live operational systems so business users can analyze and report on it without affecting the performance of those operational systems. The resulting real-time insight frees the enterprise from the constraints of working with data that no longer reflects the current status of business processes and enables it to make critical strategic decisions at the speed of business.

## About Informatica

Informatica Corporation (NASDAQ: INFA) is the world's number one independent provider of data integration software. Organizations around the world rely on Informatica for maximizing return on data to drive their top business imperatives. Worldwide, over 4,630 enterprises depend on Informatica to fully leverage their information assets residing on-premise, in the Cloud and across social networks.



Worldwide Headquarters, 100 Cardinal Way, Redwood City, CA 94063, USA Phone: 650.385.5000 Fax: 650.385.5500  
Toll-free in the US: 1.800.653.3871 [informatica.com](http://informatica.com) [linkedin.com/company/informatica](https://www.linkedin.com/company/informatica) [twitter.com/InformaticaCorp](https://twitter.com/InformaticaCorp)

© 2013 Informatica Corporation. All rights reserved. Informatica® and Put potential to work™ are trademarks or registered trademarks of Informatica Corporation in the United States and in jurisdictions throughout the world. All other company and product names may be trade names or trademarks.