Data Integration for the Real Time Enterprise

Overcoming the Challenges of Global Uncertainty

A company may be compared to a large ship that must be effectively steered at a strategic level as well as an operational level to reach its destination despite navigating through treacherous waters. Companies continually face challenging situations from competition, customer expectations, market trends, and other external constraints. Overcoming the challenges of global uncertainty requires business agility to navigate an efficient and effective course toward achieving the company’s objectives.

Companies require timely, accurate, and relevant information to add value to their products, customer relationships, and business partnerships. Business conditions in today’s world are influenced by powerful market forces such as globalization, mergers and acquisitions, regulatory compliance, fierce competition, tight operating budgets, increased demands for improved customer service, and ever-faster product delivery. The most responsive and agile companies ultimately perform the best in their industries.

The Real-Time Enterprise Enables Business Agility

For decades, organizations have depended on a host of IT systems to support operations. These systems are often specific to divisional or departmental needs. Each application or system may by itself sufficiently support specific departmental requirements. However, with the increasing complexity of market conditions, business processes, and external events, the timely interaction between these systems becomes critical to the success of the business.

The real-time Enterprise is an organization or company that has adopted business and IT solutions, along with best practices, to ensure that information on products, customers, and partners is always accurate, trusted, and available for use by critical business processes in a timely manner. The real-time Enterprise enables the business to react faster, better, and more cost effectively to evolving customer demands, changing market conditions, and approaching competitive threats.

The Information Latency Continuum

Latency requirements for delivering information cover a wide range of situations depending on specific business processes. Analytical data integration latencies typically vary from weeks to days, such as those required for certain business intelligence reports based on historical data. Operational data integration enables the delivery of information within hours, minutes, and seconds. Examples of IT projects that require operational data integration include real-time data warehouses, operational data hubs, data synchronization and replication projects, and data services or data virtualization as part of a service oriented architecture (SOA) strategy.
The success of data integration projects typically depends on the ability to meet service level agreements (SLAs) related to data latency, data completeness, and data accuracy. Following are questions to consider when planning a data integration project:

- Is the data integration project primarily for analytical or operational business purposes?
- Does the data integration project require data delivered frequently (e.g., real time or near real time) or infrequently (e.g., weekly batch windows)?
- Does the data integration project require point-of-entry data cleansing to ensure quality and avoid propagating bad data to downstream applications?
- Does the data integration project require access to a variety of data sources, including external data and advanced data transformations to ensure completeness?
- Does the data integration project require movement of large data volumes or small datasets between applications?
- Does the data integration project require ready access to the most current data available in operational systems?

**Data Integration and the Real-Time Enterprise**

Timely, trusted information is essential to sell products, service customers, and source from partners in today’s global markets. As data volumes grow and continuous data streams emerge from a variety of structured and unstructured data sources (e.g., the Web, prepackaged applications, message queues, Word, PDFs), it is becoming increasingly complex to integrate, cleanse, enrich, and deliver data to a host of consumer channels (e.g., portals, Web services, applications). Data integration is a significant competitive advantage in this arena.

Data integration requires access to any data source at any latency and of any size to be integrated, cleansed of duplicates or errors, and delivered to any consumer channel in real time or near real time. Traditional approaches (e.g., EAI, hand-coding) to data integration have proven time consuming, costly, and difficult to build and maintain. Furthermore, they don’t address issues related to quality and data governance. Following is a brief discussion of the critical capabilities to look for to support the various data integration projects in a real-time enterprise:

![The Information Latency Continuum](image)

The Information Latency Continuum describes the range of information latency requirements for analytical and operational business processes.
Real-time Data Warehousing

Loading an ODS once nightly, along with complex data aggregation to load the data warehouse, may be insufficient for large volumes of data. It is therefore much better to smooth ODS loading throughout the day. Loading an ODS in real time also provides a better recovery mechanism in case of failure. In traditional large data warehouses, if an error occurs during the ODS or data warehouse load, it is difficult to reload during the batch windows. A real-time data warehouse can be a good approach to meeting these technical and business requirements.

Data Replication

What is needed is database-independent, heterogeneous, real-time transaction replication software that’s highly scalable, reliable, and easy to use—without disrupting the performance of operational systems. It must enable IT organizations to supply the business with access to the most current data irrespective of the complexity and diversity of the IT landscape. It works in concert with the core platform to meet all the data integration, data movement, data quality, and data currency needs of an organization—such as operational reporting or operational synchronization.

Data Synchronization

An enterprise-grade data synchronization solution can enable your IT organization to achieve accurate and consistent data across operational and transactional systems. With such a solution, your IT organization can quickly build business logic that handles the most complex data synchronization projects and reuses the same logic across projects in batch, near real time, and real time. This results in high performance, greater accuracy and consistency, and a lower TCO.

Operational Intelligence

Traditional data integration only goes so far as to capture data from source systems, cleansing and transforming it and delivering it to a store. Any analysis of this data for operational decision making usually happens offline with the aid of business intelligence tools. With the combination of complex event processing and data integration, it is possible to automate the entire process of capturing data in real time in an event-driven architecture and then taking actions based on the correlation and analysis of those events. These new capabilities can help organizations and agencies respond rapidly and intelligently to emerging opportunities and threats.
Real-Time Data Quality
As soon as data is created in a system, it must be normalized and validated against standards, either automatically or through human intervention. Building data validation or data quality rules once and reusing them whenever data is created, acquired, or integrated ensures that it meets the standards of the business, while reducing the amount of effort required to continually address data quality and driving down the cost and effort of operational development efforts. Typically, this real-time information is used in conjunction with master data management platforms to deliver consolidated and reliable business-critical data to critical applications.

Multi-Enterprise Data Integration
The real-time enterprise typically requires fast onboarding of new data external data sources, flexible integration capabilities to manage a broad range of business partner technical capabilities, standardization of the onboarding process and the ability to shift it to non-development resources, and high levels of reuse to minimize redundant integration development and reduce maintenance costs. The Informatica® solution for multi-enterprise data integration provides a more centralized approach to connect, manage, monitor, and transform data with external partners, thus removing operational silos.

Operational Data Hub
Constant change to interface formats and ongoing addition of applications makes maintenance a nightmare. There are multiple points of failure. Finally, there is little control over the quality of information propagated across systems, which leads to inconsistent business processes. The data integration hub is an architecture pattern that supports internal and external information exchange between applications. The ideal solution must leverage a sophisticated data integration platform and enable self-service, auditing, and data quality. It must also integrate data at any latency.

Message-Driven Data Integration
What is needed is modern peer-to-peer messaging middleware that will support event-driven architectures for many kinds of applications, including (but not limited to) those identified as involving Big Data. Not to be confused with point-to-point messaging, peer-to-peer messaging leverages a distributed architecture similar to the “shared nothing” architectures used across the Web and in many of today’s massively scalable database systems for online transaction processing and data warehousing.

“Using Informatica’s Pushdown Optimization with Teradata allowed us to combine wireline and wireless information into an enterprise data warehouse, meeting the SLA established by our users. The daily load time for the enterprise data warehouse was decreased by 50 percent, from twelve hours to six. The ultimate result will be a unified view of our customers. Adding Informatica’s change data capture (CDC) for ADABAS helped us to deliver, within a very tight timeframe, customer information to the fraud agency. These accomplishments would not have been possible without the significant performance capabilities offered by the Informatica and Teradata platforms.”

– Vera Helena de Ávila Duarte, IT - Operations and Architecture Manager, Oi
Conclusion: Informatica for the Real-Time Enterprise

Informatica provides a highly available and scalable enterprise data integration platform for accessing, discovering, cleansing, and integrating data from virtually any business system—in any format—and delivering that data throughout the enterprise at any speed. With Informatica, enterprises can integrate and provision operational data in real time with high-performance data integration software. Informatica supports all types of business processes that require real-time or near-real-time agility, including cross-selling by customer service, point-of-sale fraud detection, real-time insurance policy quotations, SEPA and AML regulatory compliance, personalized best offers for marketing, supply chain optimization, retail out-of-stock replenishment, straight-through trade processing and reconciliation, 24x7 global operations, and many others.

With Informatica, you can support the entire data integration lifecycle for all your analytical and operational IT projects. Informatica has helped thousands of companies successfully overcome the challenges of global uncertainty to improve operational efficiency, better serve customers, improve communications with trading partners, and minimize exposure to risk. Informatica provides products and services to cost effectively access and deliver timely and trusted data in support of both analytical and operational business processes, enabling the real-time enterprise.

“Right-time data synchronization and cleansing is a key component of our migration strategy to a responsive and agile architecture. Informatica delivers right-time synchronization of trade and account data between our legacy mainframe systems and newer transactional systems to support the high performance requirements of one of the busiest options exchanges in the world. Informatica enabled us to implement the project in half the time than alternative methods.”

– Douglas M. Schafer, Jr., First Vice President
Derivatives Trading Systems Development and Back Office Enterprise Architecture Development,
Philadelphia Stock Exchange

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.