

Preparing for the Next Generation of Data Integration

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 September 2013

Table of Contents

Executive Summary	2
Your Most Critical Application... May Not Be What You Think	3
Why Do Data Integration Projects Take So Long?	4
Technical Issues	4
Business and Process Issues	5
Using Data Integration for Strategic Advantage	5
The Informatica Approach to Lean Data Integration	6
What Makes the Informatica Platform Special?	6
Conclusion	9

Executive Summary

Preparing, consolidating, and sharing data across the enterprise and with partners requires as much and possibly more time, effort, and coding than most individual business-critical applications. With the progression from the mainframe era to a distributed computing era based on integrated information, the process of extracting, cleansing, preparing, and moving data has become increasingly critical for maintaining competitive business advantage. Indeed, an organization's ability to integrate data quickly and efficiently across an ever-growing environment of distributed, interconnected systems and devices can be the key factor in its ability to compete and win.

This white paper explains why rapid data integration is now vital to competitive advantage, examines why data integration projects are so time consuming, and suggests an approach that streamlines projects to make them as much as five times faster.

Your Most Critical Application... May Not Be What You Think

Which of your company's applications demands the most time, effort, and coding effort? The first thing to spring to mind is probably a major line-of-business application, such as ERP, CRM, core banking, or claims. Now, consider what it takes to move high-quality information into and out of that application, connect it to the other applications you use internally as well as your business partners' systems, and consolidate all that data for reporting purposes. What if you added up all that time and effort, all the hours of labor expended and the lines of code written? For most modern organizations, the total would probably equal any other business-critical application—and maybe even surpass it.

Fifty years ago, a company's data all resided on a single large mainframe. Today, however, the typical information infrastructure is an abundant web of devices—on-premise and in the cloud, desktop and mobile, as large as the entire Internet and as small as tiny sensors. The more complex and distributed this web becomes, the more critical data integration has become to the organization's ability to function. The speed at which your business operates and competes is no longer limited by raw computing power governed by CPU, memory, disk, and networking. As the number of integrated endpoints has grown, the speed with which businesses compete is now limited by the speed with which they can integrate new data-generating endpoints with the rest of the data environment.

The problem is that most organizations approach data integration as a tactical activity, which makes it an inefficient series of one-off projects that are purely technological in nature. Such organizations fail to recognize that there will always be more data integration projects and that as organizations add new data sources, each subsequent project will be more complex than the last. The entire nature of computing has changed; companies with few interconnected systems today will inevitably have them tomorrow. To leverage those changes for competitive advantage, your company needs to acknowledge them and prepare a foundation of data integration that can accommodate whatever comes next.

Why Do Data Integration Projects Take So Long?

Even though IT is well aware of the trend toward more distributed systems and the need for interconnection, it often struggles with the need to deliver more data, in a more timely manner, in support of greater business agility. Each subsequent change or addition to a data source and each subsequent integration project adds another layer of complexity, making each project a little slower than the one before. The underlying reasons for this can be both technical and business oriented.

Technical Issues

One frequent technical challenge to data integration that many business users are unaware of is that information is often not stored consistently across systems. A single piece of data might be stored across a wide variety of physical environments (mainframe, database, Hadoop, cloud, etc.) and applications. Some of these, such as SaaS and partner applications, may not be under the organization's direct control. As a result, for example, a company that runs a tiered reward program might label the tiers bronze, silver, and gold in one system but 1, 2, and 3 in another. These different labels for the same information need to be reconciled before the information can be shared—and that's just one small example that, in most organizations, will have to be multiplied by thousands of attributes and data points across dozens of systems.

Another technical issue is data quality and cleanliness. Most data is dirty. Applications that allow for free-form entry of data make it easy for errors to creep into the system; even machine-generated data, such as sensor data, is error prone and must be cleaned and filtered before use. For example, a package shipping company using engine sensor data to identify potential efficiencies had to filter and cleanse the data before it could analyze the data to reduce truck idle time (and thus save millions of dollars).

These are just a few basic examples. Other technical challenges include accessing legacy data, moving large amounts of data within small time windows or small amounts of data in even smaller time windows, ensuring that all data moved is moved with 100 percent reliability, and keeping up with new technological developments in the data integration space. The result: data integration is often more technically challenging than most business and some IT people fully appreciate.

Business and Process Issues

The success or failure of data integration projects also depends on addressing certain critical business issues. One common but relatively straightforward business challenge is ensuring that different functional parts of your organization share the same definition for the same business term to avoid wasted time and later rework.

The way that business and IT communicate and work together to implement data integration projects can also make a significant difference in the time and money necessary to complete a data integration project. Informatica has found that traditional project management approaches, in which business users create large project specifications and IT then implements against those specifications, are much slower than more agile approaches with constant iteration, communication, and revision. Agile processes require a high level of collaboration and alignment between business and IT, but they are proven to be up to five times faster than the traditional approach.

Finally, many organizations make the significant error of treating each data integration project as its own custom project—starting from scratch with each one, using technologies that work in isolation, but are incompatible with each other. This approach prevents organizations from leveraging and reusing skills and tools from one project to the next. It also prevents organizations from taking advantage of advanced tooling that protects against future technological changes, such as the introduction of cloud-based integration, data virtualization, or Hadoop. To avoid this error, start from the assumption that even if your environment is not complex today, it will inevitably become more complex over time, with more endpoints and thus more need for data integration.

Using Data Integration for Strategic Advantage

Turning data integration into a strategic weapon requires organizations to treat it as a critical business process for transforming data into actionable information. That requires an evolution in data integration processes not unlike the one that transformed the auto industry from custom manufacturing in its earliest days through the production line and skills specialization that made the Model T possible, and on to today's lean manufacturing techniques that allow rapid retooling of production lines for maximum reuse. In much the same way, "lean" data integration involves these characteristics:

- Business and IT collaboration that ensures just-in-time delivery of projects
- The ability to reuse skills, knowledge, and technology across projects
- The ability to start small and grow into more sophisticated systems as integration projects mature

When put in the context of the issues discussed in this paper, these points seem obvious. Continuing to follow "custom manufacturing" techniques for data integration keeps the process slow and error prone. To evolve to a leaner process that delivers greater competitive advantage, organizations need to recognize that data integration is a multidisciplinary effort that combines people, process, and technology collaboratively across both business and IT.

The Informatica Approach to Lean Data Integration

When accurate information is critical to the success of your organization, you can't take the risk that designing your data integration processes will take too long. You can't waste time starting over from scratch every time you add a new data source or every time the technology market shifts and you need to hire people with difficult-to-find skills. Most importantly, you can't gamble that you won't get the support and services you need, when you need them, to back you up and ensure your data integration and business success. The Informatica® Platform's data integration and quality products furnish the development agility, deployment flexibility, and operational confidence required to transform raw data into actionable information.

Informatica data integration and quality products:

- Deliver **development agility** that allows our customers to complete their data integration and quality projects five times faster than alternative approaches.
- Provide **deployment flexibility** that lets you map once and deploy anywhere—to batch, real time, virtual, or even Hadoop—all from the same graphical development environment. Our ability to adopt to new technologies without requiring retraining future-proofs your investment in the Informatica Platform.
- Offer the **highest success rate** in the industry, backed up by the experience of more than 5,000 successful enterprise customers and the number one support organization for the past seven years.

What Makes the Informatica Platform Special?

Informatica customers complete their data integration jobs—from the initial definition phase all the way to deployment and management—five times faster than hand coding and open source (see Figure 1).

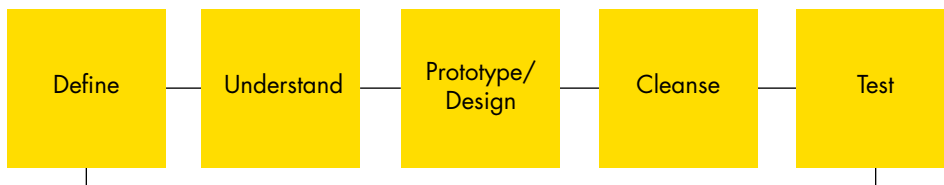


Figure 1. Development Agility: Five Times Faster.

Informatica provides a complete tool set for developing repeatable and scalable data integration and quality processes:

- A collaborative development environment allows business users and IT to **define** business terms and project definitions jointly so everyone involved starts and stays on the same page.
- Best-in-class profiling tools enable both business analysts and developers to **understand** the quality of their data and create scorecards that track improvement against a baseline.
- Fit-for-purpose development tools let business and data analysts build virtual prototypes of data integration jobs using simple Web-based tools. IT integration developers can then **convert those prototypes into production jobs** with a few mouse clicks. The iterative, agile collaboration facilitated by the Informatica Platform between business and IT significantly accelerates the development process over hand coding or open source approaches. No rewrites are necessary because the data virtualization and batch integration products are based on the same underlying Vibe™ virtual data machine.
- A combination of technology to **cleanse data** and integrated exception management workflow keeps dirty data out of the business decision-making process by cleansing it during the data movement process. IT and business users can easily collaborate in implementing an ongoing business process for continuous data quality improvement.
- Tools that automate 50 to 80 percent of **quality assurance testing** slash the overall project development cycle by 15 to 25 percent, while other tools enforce development best practices. No other vendor delivers these capabilities. It's like having a world-class data integration expert looking over the shoulder of every data integration developer in your organization.

Informatica's data infrastructure platform supports a wide variety of data integration technologies (see Figure 2) that range from free entry-level solutions to easy-to-use cloud product offerings to enterprise-class high-availability, real-time, and Hadoop-based deployments. Additionally, Informatica supplies portability across these environments so integration jobs developed to run on batch, for example, can be moved to run on Hadoop with virtually no rework required. Our technology and your organizational needs evolve together.



Figure 2. Deployment Flexibility: Map Once, Deploy Anywhere.

The ability to use a common, no-code development environment future-proofs your investment with the emergence of new integration technologies, such as Hadoop. As your integration needs change, you can change the underlying technology you use with minimal rework. For example, if your organization decides to move from on-premise applications and integration to SaaS applications, Informatica can support that move with data integration products that work both on-premise and in the cloud. Because these different products all share the same underlying Vibe virtual data machine (the integration engine used to connect, transform, and move data), changing your deployment model doesn't require you to rewrite your integration jobs.

Start small with free downloadable software or easy-to-use SaaS-based integration services, then scale in sophistication and size as your needs grow. Leverage our collaborative tools for business and IT to move from design to deployment five times faster than traditional hand coding or open source approaches. Future-proof your data integration investments with a comprehensive data infrastructure platform that delivers data integration and data quality capabilities that can span from entry-level to enterprise, on-premise to cloud, batch to real time to Hadoop—without recoding or retraining.

More than 5,000 successful customer deployments —and seven consecutive years of top ratings in customer loyalty rankings—prove that Informatica's combined data integration and data quality technology, pioneering best practices, and world-class support and services (Figure 3) provide a solid foundation for turning data into strategic advantage.



Figure 3. Operational Confidence: Proven Customer Success

Conclusion

With data sources, datatypes, data volume, and data endpoints growing all the time, the need for data integration has made it one of the most critical applications many organizations have to manage. To corral this fragmented data and create data integration and data governance processes that deliver a competitive advantage, organizations need expertise in transforming data into actionable information.

Informatica is uniquely qualified to support data integration projects ranging in scope from entry level to mission critical. Our technology is platform agnostic, allowing our customers to integrate any data with any systems. Our data infrastructure platform supports highly available systems in some of the world's most demanding environments. Key industry analysts consider us a leader in all relevant market segments, including data integration, data quality, master data management, and information lifecycle management. And we make it easy to extend the capabilities of our product family with a strong partner ecosystem with hundreds of solutions available on the Informatica Marketplace. By choosing Informatica, organizations can leverage trusted, proven technology, processes, and experience to unlock the potential in all their data—both now and well into the future.

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 to realize their information potential and drive top business imperatives. Informatica Vibe, the industry's first and only embeddable virtual data machine (VDM), powers the unique "Map Once. Deploy Anywhere." capabilities of the Informatica Platform. Worldwide, over 5,000 enterprises depend on Informatica to fully leverage their information assets from devices to mobile to social to big data residing on-premise, in the Cloud and across social networks. For more information, call +1 650-385-5000 (1-800-653-3871 in the U.S.), or visit www.informatica.com.



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 [linkedin.com/company/informatica](https://www.linkedin.com/company/informatica) 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.