

# Jump-Start Your Data Warehousing Initiatives

*Using the Informatica Platform and Informatica Velocity  
to Guide Your Project Success*

WHITE PAPER



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 October 2008.

## Table of Contents

|   |           |
|---|-----------|
| <b>Executive Summary</b> . . . . .  | <b>2</b>  |
| <b>Rethink Your Approach to Data Warehousing.</b> . . . . .   | <b>3</b>  |
| <b>Build Your Data Warehousing Initiatives on a Comprehensive, Open, and Unified Data Integration Platform.</b> . . . . . | <b>4</b>  |
| Broad Connectivity and Flexible Delivery . . . . .  | 5         |
| Cross-Functional Collaboration Between Business and IT . . . . .  | 5         |
| Global Metadata Infrastructure . . . . .  | 5         |
| Data Quality Management. . . . .  | 6         |
| Team-Based, Visual Development Environment. . . . .   | 6         |
| Superior Performance and Scalability . . . . .  | 6         |
| <b>Speed Time to Value with Informatica Velocity</b> . . . . .  | <b>7</b>  |
| More Than Just an Implementation Methodology . . . . .  | 7         |
| <b>Guide Data Warehousing Initiatives to Success with Informatica Velocity</b> . . . . .                                  | <b>9</b>  |
| Start Your Project Right . . . . .  | 9         |
| Fine-Tune Your Game Plan . . . . .  | 10        |
| Ensure Executive Support and Organizational Alignment . . . . .   | 11        |
| Expand Incrementally Toward Broader Deployment . . . . .  | 12        |
| Complement Your Existing Methodology . . . . .  | 12        |
| <b>Put Data Integration Best Practices to Work in Your Data Warehousing Environment.</b> . . . . .                        | <b>13</b> |
| Key Management in Data Warehousing Solutions . . . . .  | 13        |
| Building Data Audit/Balancing Processes. . . . .  | 14        |
| Effective Data Standardizing Techniques. . . . .  | 14        |
| Designing Data Integration Architectures . . . . .  | 14        |
| Managing Internal and External Reference Data . . . . .   | 14        |
| Error Handling Strategies. . . . .  | 15        |
| Error Handling Processes. . . . .   | 15        |
| Creating Inventories of Reusable Objects and Mappings . . . . .   | 15        |
| Metadata Reporting and Sharing. . . . .   | 15        |
| Recommended Performance Tuning Procedures . . . . .   | 16        |
| <b>Extend the Reach of Your Data Warehousing Projects</b> . . . . .   | <b>16</b> |
| Informatica Velocity for Enterprise Integration Initiatives . . . . .   | 16        |
| Additional Project Resources . . . . .  | 17        |
| <b>Conclusion.</b> . . . . .  | <b>17</b> |

## Executive Summary

Data warehousing is a strategic initiative whose goal is to reduce costs, increase productivity, and drive competitive advantages by managing data. Data warehousing supports business intelligence (BI), customer relationship management (CRM), enterprise resource planning (ERP), and many other applications. Data warehousing projects can deliver big results, but they can also pose big risks, particularly when it comes to implementation.

How can your IT organization mitigate implementation risks? Can you use your existing hand-coded or point solutions to meet business demands? What methodologies do you follow to ensure project success and meet or exceed your goals? Is there a way your IT organization can proceed incrementally while establishing a long-term foundation for data warehousing?

As a data warehousing professional, you play a critical role in selecting the right technology solutions and leveraging best practices that ensure your data warehousing projects are deployed successfully and on time. You can champion important goals, such as accelerating the launch of the project, increasing the number of supported projects, and getting greater value from your data warehousing projects.

In this white paper, you'll learn how to:

- Rethink your approach to data warehousing
- Build your data warehousing initiatives on a comprehensive, open, and unified data integration platform
- Speed time to value with Informatica® Velocity<sup>SM</sup> methodology
- Jump-start your data warehousing initiatives
- Put data integration best practices to work in your data warehousing environment
- Extend the reach of your data warehousing projects with additional resources from Informatica

## Rethink Your Approach to Data Warehousing

Data warehousing projects can deliver big results, but they can also pose big risks. Many data warehousing projects go over budget or schedule, and some wind up being cancelled altogether. To minimize these risks, your IT organization may be considering an incremental approach. You could start by delivering smaller departmental data marts that focus on the needs of individual business constituencies, such as adding data fields to your existing business intelligence practice. In turn, these data marts would feed into a data warehouse that serves multiple subject areas.

Under pressure to deliver data marts and data warehouses more cost-effectively and quickly, many IT organizations still use hand-coding or ad hoc assortments of disparate tools to accelerate implementation. Although these tools are readily available and inexpensive, in the long run they create a system that is difficult to maintain and not easily adaptable to changing business requirements.

What's the alternative?

Modern extract-transform-load (ETL) and data integration techniques have become industry best practices for building cost-effective, reliable data marts and data warehouses. A tools-based data integration approach has been proven to reduce data warehousing project costs by an average of 35 percent and to speed their delivery.<sup>1</sup>

Your IT organization needs a comprehensive solution to help you deliver data marts and data warehouses quickly and economically. This solution should free you from the expense, wasted time, and maintenance headaches of hand-coded solutions or ad hoc arrangements of different tools.

Informatica offers a complete data warehousing solution to help your IT organization jump-start its data warehousing initiatives and ensure that they are completed on time, on budget, and within scope. This solution consists of the industry-leading **Informatica data integration platform** and **Informatica Velocity**, our proven implementation methodology, and is supported by Informatica's world-class partner ecosystem and a large, active developer network.

<sup>1</sup>Source: Forrester Research, "The Total Economic Impact of Deploying Informatica PowerCenter," 2004.

## Build Your Data Warehousing Initiatives on a Comprehensive, Open, and Unified Data Integration Platform

Instead of approaching data warehousing initiatives using cumbersome, labor-intensive hand-coding or multiple-point solutions, many IT organizations recognize the value of using a comprehensive, open, and unified data integration platform. A platform-based approach offers many advantages, including:

- **Greater cost savings.** Save maintenance and support costs over time versus save license costs.
- **Better infrastructure.** Establish a new foundation versus build on existing architecture.
- **Physical environmental readiness.** Migrate to a future environment versus keep what works.
- **Shared knowledge base.** Benefit from the ecosystem and influence technical directions versus sustain internal technology and resources.
- **Implementation expertise.** Build a competency center versus implement project by project.
- **Time to market.** Deliver more projects faster versus manage current project load.

The Informatica data integration platform is the ideal foundation for today's data warehousing projects. Our platform consists of:

- **Informatica PowerExchange®** for accessing data and managing connectivity
- **Informatica Data Explorer™** and **Informatica Data Quality™** for managing data quality
- **Informatica PowerCenter®** for integrating and delivering data

The platform delivers on the six major capabilities your IT organization needs to jump-start your data warehousing projects:

1. Broad connectivity and flexible delivery
2. Cross-functional collaboration between business and IT
3. Global metadata infrastructure
4. Data quality management
5. Team-based, visual development environment
6. Superior performance and scalability

## Broad Connectivity and Flexible Delivery

Your data warehouse needs to be able to access all your enterprise data. With the universal data access capabilities of Informatica PowerExchange, your IT organization can achieve broad connectivity and prebuilt access to:

- Major enterprise and packaged applications, whether on-premise, outsourced, or hosted software-as-a-service
- All major enterprise database systems and data warehousing environments
- Mainframe systems
- Midrange systems
- Message-oriented middleware
- Industry-wide technology standards, such as email, LDAP, and Web services

Whether your team needs real-time, batch, or changed data, PowerExchange provides a unified environment for accessing all data with lower cost and complexity.

## Cross-Functional Collaboration Between Business and IT

Informatica PowerCenter enables global teams of developers, analysts, and administrators to work faster and better together. Data warehousing projects can be readily administered and managed even when team members have different roles and responsibilities, or when they're located in different offices—whether onshore or offshore.

Two key features of PowerCenter increase collaboration across your data warehousing and business intelligence teams:

1. **Mapping Architect for Visio.** With this PowerCenter feature, developers can accelerate and standardize design using a data-flow design tool based on Microsoft Visio and mapping templates incorporating Informatica Velocity best practices.
2. **Mapping Analyst for Excel.** With this PowerCenter feature, business analysts can create mapping specifications in Microsoft Excel and then generate PowerCenter data integration mappings from those specifications to jump-start development.

## Global Metadata Infrastructure

The Informatica platform's powerful metadata analysis and management capabilities help ensure visibility and auditability for your data warehousing environment.

**Metadata Manager**, one of the key features of PowerCenter Advanced Edition, aggregates and links metadata from a wide variety of data sources and third-party tools in an integration metadata catalog. Metadata Manager also supplies information about structure, end-to-end impact analysis, and report-to-source data lineage, not readily available in a hand-coded environment.

## Data Quality Management

If your IT department has hand-coded cleansing/matching algorithms, you know it's often difficult to deliver data that the business can trust. If your data warehouse is going to be successful, it needs to be populated and maintained with complete, consistent, high-quality data. With **Informatica Data Explorer** and **Informatica Data Quality**, your IT organization can approach data quality as an ongoing, enterprise-wide process.

Informatica Data Explorer delivers a complete and completely accurate picture of the content, quality, and structure of enterprise data. Combining robust data profiling and mapping capabilities, Informatica Data Explorer helps your team investigate, document, and resolve data quality issues.

Once those issues have been identified, your team can use Informatica Data Quality, with its powerful data analysis, cleansing, matching, reporting, and monitoring capabilities, to remediate and enhance the accuracy and value of your data assets. You can implement, manage, and monitor enterprise-wide data quality using executive dashboards and real-time notifications, which are becoming standard best practices.

## Team-Based, Visual Development Environment

The Informatica platform provides your IT team with an easy-to-use visual development environment and centralized administration to increase your productivity and ensure that your data warehouse is brought on line quickly.

Defining transformations is a simple “drag and drop” process. With the platform's **Administration Console**, you can centralize all configuration and monitoring tasks to ensure that development work can be shared and reused across platforms and projects without having to recode. This helps your IT organization respond more quickly to changing business needs while holding down development costs.

## Superior Performance and Scalability

Data warehousing initiatives increasingly require high levels of performance and scalability to meet the demands of large data volumes, complex processing, and reduced latency requirements. PowerCenter delivers superior throughput and 64-bit processing. Performance and scalability are enhanced through capabilities in:

- **High availability and seamless failover and recovery.** By ensuring seamless failover and recovery of all PowerCenter components, the PowerCenter High Availability Option minimizes service interruption in the event of a hardware or software outage. Your IT organization can cost-effectively scale to meet increased data demand, save hardware costs, and reduce the expense and risks associated with data downtime.
- **Grid computing.** The PowerCenter Enterprise Grid Option provides scalability within a grid computing environment, while reducing the administrative overhead of supporting a grid. Your IT organization can automatically balance loads in response to run-time changes in data volumes or utilization rates.
- **Pushdown optimization.** The PowerCenter Pushdown Optimization Option lets you push data transformation processing to the resource offering the greatest performance advantage. Your IT organization saves hardware costs, enhances productivity, and increases system performance.

## Speed Time to Value with Informatica Velocity

A change in your data warehousing approach creates challenges and risks. Moving from hand-coding or using ad hoc assortments of disparate tools to a platform-based approach can offer clear advantages. But the change in approach can add complexity to pre-existing and often already complex data warehousing environments. New activities must be managed while still executing day-to-day operations. New upfront costs must be accounted for.

Your IT organization needs expert guidance to diffuse or avoid these risks altogether. You need field-tested tips, best-practice techniques, and comprehensive tracking capabilities to keep your data warehousing projects within scope, on schedule, and on budget.

### More Than Just an Implementation Methodology

Based on thousands of successful, real-world deployments, Informatica has developed an implementation methodology called Informatica Velocity.

Much more than just another project methodology, Velocity provides your IT organization with intuitive, practical task breakdowns, project structures, and data warehousing best practices. These field-proven best practices are continuously updated and refined to help you deliver your data warehousing projects while mitigating risks and overruns. Leveraged in conjunction with formal Informatica product training, Velocity helps your team increase its productivity immediately.

Informatica Velocity is based on 10 years of experience implementing hundreds of data warehousing projects. The methodology represents the collective knowledge of experienced consultants who have worked with hundreds of organizations to plan, develop, deploy, and maintain successful projects. Reflecting a wide range of system environments and user needs, Velocity guides your project team through the entire project life cycle so that you avoid common pitfalls and achieve predictable, timely results.

Velocity furnishes detailed, project-specific guidance through each data warehousing project phase:

- **Analyze.** Each project begins with planning based on a thorough understanding of business requirements.
- **Architect.** The analysis serves as the basis for defining the technical architecture and environmental strategy.
- **Design-Build-Test-Deploy.** Each phased iteration supplies part of the overall solution until the solution is complete.
- **Operate.** The system faces new challenges as it moves into day-to-day operation.
- **Manage.** Continuous tracking and measurement activities span the full project life cycle.

Figure 1 shows how Velocity provides guidance throughout the entire life cycle of a data warehousing project.

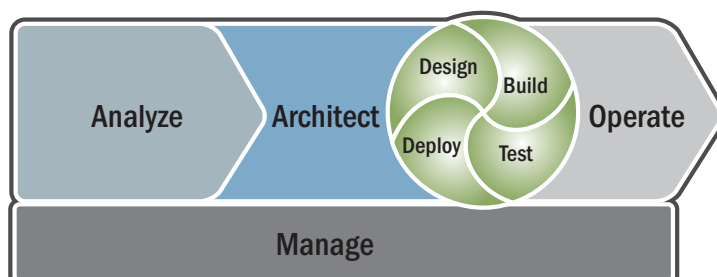


Figure 1. Informatica Velocity provides detailed guidance through each project phase for complete project life-cycle support.

“With Informatica Professional Services and the Velocity methodology, the results have been one successful deployment after another with no re-work. Thanks to the expertise of Informatica’s consultants, their methodology, and their best practices, all of our data integration projects get done right the first time.”

— Donna Capraro  
Data Warehouse Manager  
University of California, Los Angeles

In addition, Velocity delivers all the elements your IT organization needs to guide successful data warehousing implementations, including these:

- **Work breakdown structures** outline all activities—such as project phases, subphases, and tasks.
- **Role definitions** help team members focus on the appropriate tasks for their function.
- **Data warehousing best practices** provide guidance by technology type (e.g., databases, applications).
- **Sample deliverables** furnish practical examples of requirement specifications, test plans, and development schedules.

Informatica Velocity supplies sample deliverables for data warehousing projects. These documents provide a quick framework for IT teams to use when documenting specific areas of the data warehousing project. Sample deliverables include:

- Business Requirements Specification
- Change Request Form
- Data Migration Communication Plan
- Data Quality Plan Design
- Database Sizing Model
- Functional Requirements Specification
- Information Requirements Specification
- Issues Tracking
- Mapping Inventory
- Mapping Specifications
- Metadata Inventory
- Migration Request Checklist
- Project Plan
- Project Roadmap
- Project Role Matrix
- Prototype Feedback
- Restartability Matrix
- Scope Change Assessment
- Source Availability Matrix
- System Test Plan
- Target-Source Matrix
- Technology Evaluation Checklist
- Test Case List
- Test Condition Results
- Unit Test Plan
- Work Breakdown Structure

## Guide Data Warehousing Initiatives to Success with Informatica Velocity

Data warehouses are often built by IT teams that have been redeployed from other implementation efforts, such as system implementations and upgrades. For these team members, implementing a data warehouse can be quite a change from their past IT work. This section explains how your IT team can use Informatica Velocity to:

- Start your data warehousing project right to ensure high-quality delivery
- Fine-tune your game plan
- Ensure executive support and organizational alignment
- Expand incrementally toward broader deployment
- Complement your existing methodology

### Start Your Project Right

When embarking on a data warehousing project, it's important to recognize the differences between a data warehousing project and a typical system implementation. Here are some characteristics of a data warehousing implementation:

- Data sources are typically from many disparate systems located both internally and externally.
- Data models are used to understand relationships and business rules within the data.
- Data volumes for both data integration and analytic reporting are high.
- Historical data is maintained, often for years.
- Data frequently is stored as both detailed-level data and summarized or aggregated data.
- The underlying database system is tuned for querying large volumes of data rather than for inserting single transaction data.
- Data warehouse data supports tactical and strategic decision making, rather than operational processing.
- A successful data warehouse is driven by the information needs of the business.

Figure 2 shows the characteristics of a data warehousing project.

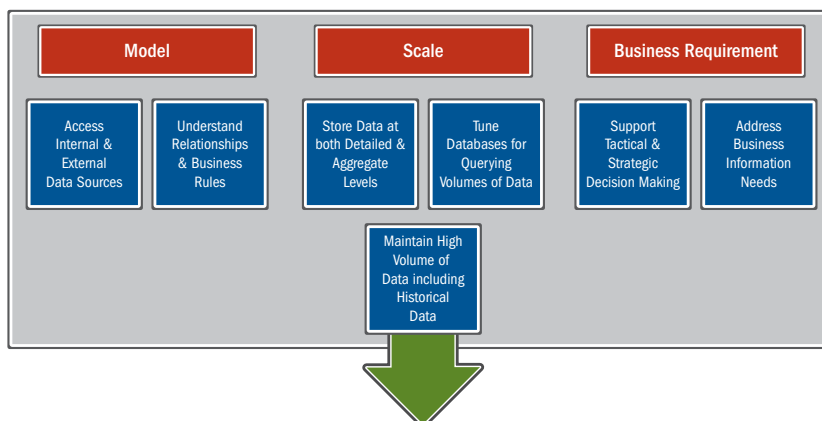


Figure 2. It's important to take into account the complexities of a data warehousing project—including the model, scale, and business requirements—from the onset.

Informatica Velocity can help your IT team get off to a strong start on your data warehousing project. Velocity can guide your team in determining your project's scale and business requirements. Based on these requirements, you model, structure, and populate data in a relational database. You design the data integration environment based on available operational application sources. You pull, cleanse, and transform data from these sources to populate a subject area database. Once the subject area database contains the data, you can supply BI reports to end users. You can also identify detailed reporting requirements and design corresponding BI data marts that capture all the specific reporting facts and dimensions.

## **Fine-Tune Your Game Plan**

Not every data warehouse project is a new implementation. Often data warehouses are deployed in phases where new subject areas, new data sources, or enhanced reporting are added to the existing solution. Velocity can help your IT team fine-tune your game plan depending on the type of implementation you're undertaking.

### **New Business Data Project: Avoid Costly Rework and Ensure User Acceptance**

This type of project addresses the need to gather data from an area of the enterprise that is not familiar with the business data or requirements. Logical data modeling is a crucial step in this type of project because you need to thoroughly understand and model data requirements from a business perspective. The logical data model serves as the blueprint for all follow-on project work.

Velocity guides your team in gathering business requirements and constructing the logical data model. Providing sample deliverables, Velocity also helps you model, design, and implement new dimensional data marts and BI reporting offerings. Velocity enables your IT team to meet end users' information and access requirements in the testing phase, avoiding costly rework.

### **Enhanced Data Source Project: Promote Auditability and Visibility**

This type of project addresses the need to add a new data source or to alter an existing data source, but always within the context of logical data structures and definitions that are already established. No logical data modeling is necessary because there are no new business data requirements. Minor adjustments to the physical model and database may be needed to accommodate changes in data volumes. Volumes increase when new sources are added, or new or altered views may be required to report on new data instances that may now be available to end users.

Velocity helps you make these adjustments and improve business insight from source to target. When enhancing existing data, it's important to manage metadata to track data from the physical data sources through the data integration process and to business intelligence data marts and reports. Velocity can guide your IT team in developing a metadata strategy that provides end-to-end data lineage and impact analysis so that you can ensure your data is auditable and visible from source to target.

## Enhanced Business Intelligence Requirements Project: Respond to Business Demands Quickly

This type of project focuses exclusively on expanding or altering the BI reporting and query capability using existing subject area data. New or altered data (in structure or content) does not need to be added to the warehouse subject area database. New or altered dimensional data mart tables or views may be required to support the BI enhancements; otherwise the majority, if not all, of the work is within BI. In this case, Velocity can improve the productivity of developers and analysts as they change the data for BI enhancements.

## Ensure Executive Support and Organizational Alignment

Successful data warehousing projects are backed by a strong organizational commitment to enterprise analytics. The value and return on investment (ROI) must be clearly articulated early in the project. The costs and time needed to achieve results must be fully explored and understood. Your IT organization should assess the business case and determine executive sponsorship early on in the project.

Velocity provides documentation templates and guidelines to help you gather requirements and develop the business case. Velocity can also help you set up your project team. As **Figure 3** shows, Velocity supplies sample project team organizational charts. The methodology identifies typical project team roles and responsibilities for both part- and full-time team members. Velocity also provides a role description for each title that you can use to recruit new team members or adjust the roles of existing team members.

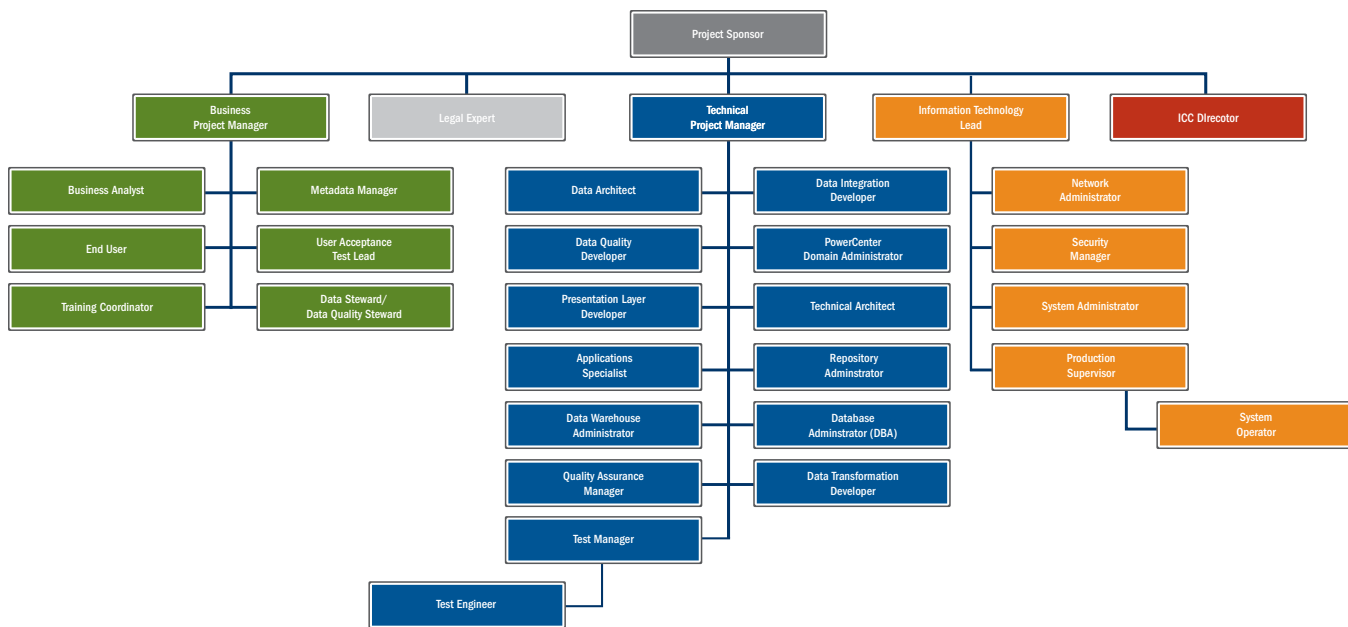


Figure 3. Informatica Velocity provides sample project team organizational charts to help you establish your data warehousing project teams.

## Expand Incrementally Toward Broader Deployment

Often data warehousing initiatives strive to achieve a “single source of truth” across the entire enterprise and across all data stores. Delivering this in a “big bang” approach nearly always fails. By the time all the data modeling, rationalization, and integration have taken place across the enterprise, the value of the project is called into question, and the project is at risk of being delayed or cancelled.

It’s important for project teams to find areas of high business value and then deliver on those areas quickly. Velocity helps your IT team sustain project momentum by demonstrating results incrementally. For instance, your team may start out with data warehousing for financial reporting and analysis and then expand this effort into human resources and planning. Meeting regular success milestones and demonstrating incremental business value keeps the executive sponsorship engaged and proves the value of the data warehouse to the organization early and often. With Velocity, you can plan deployment phases to deliver value quickly and continuously throughout the project life cycle.

## Complement Your Existing Methodology

Velocity is aligned with the standard software development life cycle (SDLC). As such, you can easily adopt it to support your existing methodology, as Figure 4 illustrates.

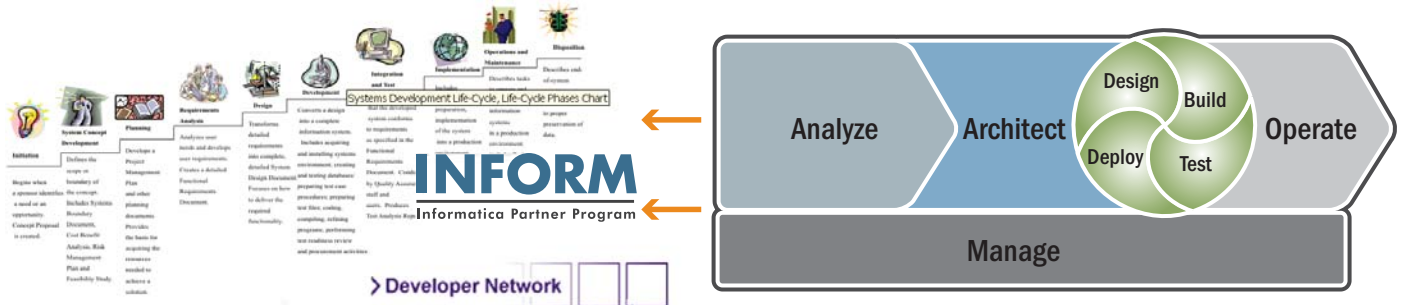


Figure 4. Informatica Velocity complements your existing software development life cycle.

Furthermore, Velocity is open to Informatica partners and the Informatica Developer Network (IDN), the global ecosystem of the Informatica development community. Because of its alignment with the SDLC, the entire community benefits from Velocity while leveraging and extending their existing methodologies. In addition, Velocity also provides detailed product-specific tips and techniques to speed your implementation.

## Put Data Integration Best Practices to Work in Your Data Warehousing Environment

Informatica Velocity provides detailed, field-proven data integration best practices. This collection of helpful documents can guide your IT team in using Informatica products to implement data warehousing projects while mitigating risks and overruns. These best practices represent more than 10 years of Informatica's hands-on data warehousing experience.

This section summarizes Informatica Velocity's top ten data warehousing best practices:

1. Key Management in Data Warehousing Solutions
2. Building Data Audit/Balancing Processes
3. Effective Data Standardizing Techniques
4. Designing Data Integration Architectures
5. Managing Internal and External Reference Data
6. Error Handling Strategies
7. Error Handling Processes
8. Creating Inventories of Reusable Objects and Mappings
9. Metadata Reporting and Sharing
10. Recommended Performance Tuning Procedures

Informatica's top ten best practices for data warehousing are described at a high level in this section. The full descriptions of each best practice are available in the *Informatica Velocity Guide*, which is available to Informatica customers and partners.

### Key Management in Data Warehousing Solutions

Key management refers to the techniques used to manage key allocation in a decision support RDBMS to create a single view of reference data from multiple sources. Key management involves 1) key merging/matching, 2) handling missing keys, and 3) handling unknown keys.

All three techniques apply to a reference data store, whereas only the missing and unknown keys are relevant for an operational data store. Informatica recommends a key management approach that ensures loading everything extracted from a source system into the data warehouse. This approach represents a departure from many traditional data warehouse solutions that apply logical and data warehouse (surrogate) key strategies when errors are loaded and transactions rejected from referential integrity issues.

Key management should be handled at the data integration level, so that it is transparent to the BI layer with a single view of key allocation. The major advantage of this approach is that any aggregate values derived from the transaction table will be correct because the transaction exists in the data warehouse rather than in some external error processing file waiting to be fixed.

## Building Data Audit/Balancing Processes

Verifying that the data in an application is complete can be challenging. It's also difficult to confirm that all the appropriate data was extracted from a source system and propagated to its final target.

Establishing an efficient and reliable data audit/balancing process can help your IT team increase its productivity. Such a process is particularly important for businesses that are either highly regulated internally or that have to comply with a host of external government regulations, such as Sarbanes-Oxley, Basel II, HIPAA, and the United States Patriot Act.

When setting up a data audit/balancing process, your IT team should produce a set of common tables to house various control metrics regarding the data integration process. Ultimately, business intelligence reports furnish insight at a glance to verify that the correct data has been pulled from the source and completely loaded to the target.

## Effective Data Standardizing Techniques

Your IT team should streamline data cleansing and standardization processes (or plans) to shorten development timelines. Data cleansing refers to the removal of irrelevant information and “noise” from data, such as:

- Personal names
- “Care of” information
- Excess character spaces
- Punctuation from postal address

Data standardization refers to modifying the appearance of data so that it takes on a more uniform structure and is enriched by deriving additional details from existing content.

## Designing Data Integration Architectures

A sound data integration architecture is crucial for the success of data warehousing projects. Historically, IT organizations have approached the development of a data warehouse or data mart as a departmental effort—not as an enterprise-wide initiative. The result has been silos of corporate data and analysis that often contradicts itself.

Your IT team should implement a data integration architectural foundation to ensure that your data warehousing projects can evolve and scale to meet changing business needs over time. The proper architecture can isolate the application component (business context) of the data warehousing solution from the technology. Allowing for the reuse of skills, design objects, and knowledge, an enterprise-wide data integration architecture can help your IT team readily extend data warehousing efforts to meet future business demands.

## Managing Internal and External Reference Data

Your IT team should develop and manage reference data sources that can be used with data quality plans. The goal is to use reference data files and data quality plans to ensure the smooth transition from development to production.

Reference data files can be used to verify or enhance the accuracy of the data in the data quality plans. A reference data file is a list of verified correct terms and their acceptable variants. It may be a list of employees, package measurements, or valid postal addresses—any data set that provides an objective reference against which project data sources can be checked or corrected.

## Error Handling Strategies

For your data warehousing project to be successful, it must deliver timely, accurate data to end users. It's imperative that data in the warehouse be as current as possible because the most current data is the most relevant to business operations.

Transactional systems can function even with a certain amount of error. The impact of an individual transaction error on business operations, as a whole, is usually minor. Corrections can be made to erroneous data after the error has been identified.

In data warehouse systems, however, any error (e.g., for a particular load instance) not only affects a larger number of data items, but also may potentially distort key reporting metrics. Erroneous data cannot be left in the warehouse "until someone notices" because such information may drive business decisions. It's important for your IT team to proactively manage errors, identifying them before or as they occur. If errors happen, it is equally important either to prevent them from getting to the warehouse at all or to remove them from the warehouse immediately.

## Error Handling Processes

When implementing a process for handling errors, your IT team should consider: 1) error identification, 2) error retrieval, and 3) error correction.

This best practice describes how your IT team can take advantage of best-of-breed error management technology available in Informatica PowerCenter to execute error handling processes. This technology includes:

- Relational database error logging
- Email notification of workflow failures
- Session error thresholds
- Robust reporting capabilities
- Data profiling

## Creating Inventories of Reusable Objects and Mappings

Your IT team should identify and create inventories of reusable objects and mappings based on business rules to increase the team's productivity. Create these inventories in a folder, or create shortcuts across folders (local shortcuts) or shortcuts across repositories (global shortcuts).

Before creating an inventory of reusable objects or shortcut objects, be sure to review the business requirements and look for any common routines or modules that may appear in more than one data movement.

## Metadata Reporting and Sharing

Your IT team should develop and implement a strategy for reporting on and sharing metadata. Detailed information or metadata comments can be entered for all repository objects (e.g., mapping, sources, targets, transformations, ports). Also, information about column size and scale, datatypes, and primary keys should be stored in the repository.

The decision on how much metadata to create is often driven by project timelines. While it may be beneficial for a developer to enter detailed descriptions of each column, expression, or variable, doing so requires extra time and effort. But once that information is fed into the repository, anyone on the team can retrieve it at any time.

## Recommended Performance Tuning Procedures

When a session or workflow is not performing at the expected or desired speed, your IT team should consider performance tuning to diagnose the problems that may be adversely affecting the data integration architecture. Although the Informatica platform has its own performance settings that can be tuned, you must consider your entire data integration architecture, including the UNIX/Windows servers, network, disk array, and the source and target databases to achieve optimal performance.

Often an external issue is the cause of the performance problem. To determine the cause, you need to execute the performance tuning steps in a specific order. Doing so enables you to methodically rule out individual pieces and narrow down the specific areas on which to focus your tuning efforts.

## Extend the Reach of Your Data Warehousing Projects

When your IT team is ready to extend your data warehousing projects across the enterprise, Informatica has resources to support you.

## Informatica Velocity for Enterprise Integration Initiatives

Informatica Velocity can guide your IT organization through the project-level tasks of deploying a data warehouse. As **Figure 5** shows, Velocity can also guide your team as you extend your data warehousing initiatives across other enterprise-wide initiatives—such as data governance programs, service-oriented architectures, and Integration Competency Centers. To the extent that one or all of these may exist in your organization, these types of initiatives will ultimately impact the method and architecture for your data warehousing project.

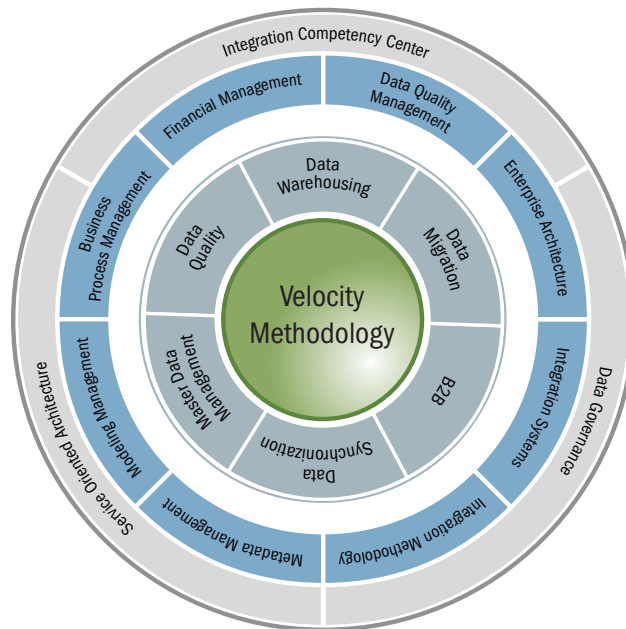


Figure 5. Velocity helps you extend your data warehousing initiatives across other enterprise-wide initiatives.

## Additional Project Resources

When starting or revamping your data warehousing practice, you need the sure hands of experienced and savvy resources to minimize costs and implementation risk. Informatica has forged robust partnerships with many of the foremost global system integrators and technology and services providers to deliver innovative data warehousing solutions. In addition, the Informatica Developer Network (IDN) has 43,000+ worldwide members who are dedicated to providing developers with data integration best practices and proven techniques. IDN is an active community of developers that shares technical tips and tricks on line and at user group meetings.

## Conclusion

Informatica offers a complete data warehousing solution to help your IT organization jump-start its data warehousing initiatives and ensure that they are completed on time, on budget, and within scope. This solution consists of the industry-leading Informatica data integration platform and Informatica Velocity, our proven implementation methodology, and is supported by Informatica's world-class partner ecosystem and a large, active developer network.

With the Informatica data warehousing solution, your IT organization can:

- Accelerate the development and deployment of data marts and data warehouses
- Lower the costs of implementing and maintaining your data marts and data warehouses
- Establish a robust, unified foundation for extending your data warehousing practice enterprise-wide

Informatica can help you put leading best practices in modern ETL and data integration to work in your environment so you can readily reuse development work from one project to the next. Informatica helps your IT organization do more with less.

## Learn More

Learn more about how Informatica can help you jump-start your data warehousing initiatives. Find out more at [www.informatica.com/solutions/enterprise\\_data\\_warehouse](http://www.informatica.com/solutions/enterprise_data_warehouse) or call (800) 653-3871. For more information about the *Informatica Velocity Guide*, contact [methodology@informatica.com](mailto:methodology@informatica.com).

## About Informatica

Informatica is a leading provider of enterprise data integration software and services. With Informatica, organizations can gain greater business value by integrating all their information assets from across the enterprise. Thousands of companies worldwide rely on Informatica to reduce the cost and expedite the time to address data integration needs of any complexity and scale.



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 [www.informatica.com](http://www.informatica.com)

Informatica Offices Around The Globe: Australia · Belgium · Canada · China · France · Germany · Japan · Korea · the Netherlands · Singapore · Switzerland · United Kingdom · USA

© 2008 Informatica Corporation. All rights reserved. Printed in the U.S.A. Informatica, the Informatica logo, and The Data Integration Company 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 of their respective owners.

6903 (10/02/2008)