Accelerate Application Data Migration with Informatica
Table Of Contents

Executive Summary ........................................................................... 1

Data Migration Planning Guarantees
Cost-effective Deployment .......................................................... 1

Data Migration in Depth ............................................................... 2

Selecting the Ideal Solution ......................................................... 3

Data Migration with Custom Code ......................................................... 4
Data Migration with Application Vendor's Solution ............................... 4
The Informatica Approach ................................................................. 4

The Informatica Data Integration Platform:
The Complete Solution ............................................................... 5

Metadata-centric Architecture Enhances the Value of Information ...... 5
  Reusability Accelerates Development .................................................. 5
  Impact Analysis and Data Lineage Provide Visibility .................................. 6
  Operational Metadata Maximizes Efficiency ......................................... 6

Flexibility Enables Migration in Batch or Real Time .............................. 6
Build-in Data Quality Helps Ensure User Adoption and Acceptance ........ 7
Iterative Methodology Accelerates Development ..................................... 7
Build-in Transformations and Ease-of-use Shorten Lead Times .......... 8
Open, Flexible Architecture Leverages Existing Investments ............... 9
Summary ................................................................................... 10
Executive Summary
Integration is now the most critical “high-ticket” item in enterprise IT budgets. The compelling business case for a single, global view of customers, products, and company financials is universally accepted, but the challenges to achieving this goal are formidable.

In today’s rapidly evolving business and regulatory environment, organizations must connect an increasing number of disparate legacy, packaged, and custom applications into a seamless information infrastructure to:
• Support changing and evolving business processes
• Support new regulatory requirements such as Sarbanes-Oxley, Basel II, and the Patriot Act
• Provide accountability through operational transparency
• React more quickly to competitive and market pressures
• Reduce IT costs

This solution brief outlines the key role of data migration in ERP instance consolidations, new packaged or custom application implementations, and the upgrade process for existing application installations. Specific guidelines for selecting optimal data migration and conversion solutions are offered, and the unique features and benefits of the Informatica solution are presented.

Data Migration Planning Guarantees
Cost-effective Deployment
The data migration phase of an application implementation, upgrade, or instance consolidation project can extend up to several years and requires rigorous project planning. Detailed project planning is required because enterprises have traditionally underestimated the challenges involved in the data migration process, including the high cost of system maintenance and administration, the cost of developing and documenting custom interface tables, and addressing complex requirements including data quality.

Many organizations learn—often mid-project— that the data migration phase of an application implementation, upgrade, or instance consolidation can consume up to 40 percent of the project’s entire budget. Industry research has shown that the average data migration project comes in over 10 times the budgeted amount.
Because data migration is such a complex process, it is important to view the data migration component of a new application implementation, upgrade, or instance consolidation as an entirely separate project. The ideal data migration project plan is dependent on a number of factors, including:

- Incorporating best-of-breed data integration tools and methodologies
- Taking into consideration the full application lifecycle, including upgrades and synchronization
- Understanding the requirements of the business, then setting and communicating clear project expectations
- Carefully selecting project leaders, roles, and responsibilities

Data Migration in Depth

Data migration is complicated because of the need to transform data from a wide variety of transactional, legacy, relational, third-party, real-time, and historical data sources into a new schema and supporting data types. This requires data to be:

- **Profiled and extracted** from current systems, applications, databases, and files
- **Cleansed** of incorrect, redundant, and outdated records
- **Transformed** to support new data types, definitions, taxonomies, and schemas
- **Tested and validated** to ensure migration routines perform according to defined specifications
- **Loaded** into the new application environment

These vital processes are further complicated due to the large number of disparate systems that need to be tied together to support application implementations or instance consolidation projects. According to a recent survey of more than 350 firms, the typical organization relies on more than 50 core line of business applications, and companies with more than $1 billion annual revenue have as many as 500 of these systems.

Moreover, once enterprises deploy new applications, they must continually upgrade them to remain competitive and receive maximum return on investment. But, the accelerating cycle of application upgrades, while delivering important new functionality, presents an entirely new set of integration challenges. Every upgrade is a major IT initiative—one that may easily take a full year and up to 30 percent of the original purchase and installation cost to complete. Upgrades often mandate fundamental changes in IT infrastructures and business practices\(^1\) as well.

\(^1\) CIO Magazine. *Enterprise Software Upgrades—Less Pain, More Gain*, November 15, 2002
In fact, a recent AMR research survey found the testing, data conversion, and migration components of an ERP upgrade to be the most difficult.\(^2\) Exacerbating this challenge is the fact that enterprise software upgrades are occurring more frequently. The average time between major upgrades has shrunk from three years in the early 1990s to 18 to 24 months.\(^3\)

Since data migration is such a vital, costly, and resource intensive phase of both initial deployments and long-term application maintenance, enterprises should evaluate only those data migration products offering a proven solution, with a flexible architecture, methodology, and tool set capable of supporting both immediate and long-term requirements.

### Selecting the Ideal Solution

The first step in evaluating data migration solutions is to determine the scope of the overall effort. It is useful to review the following factors:

- Number and complexity of legacy systems
- Documentation requirements
- Type of migration (i.e., batch, one-time, synchronous, or real time)
- Cost estimation for the full project lifecycle (i.e., upgrades, maintenance, customization, and training)
- Legacy data quality
- Amount of data and history to be converted
- Target application architecture (i.e., schema, tables, data types)
- Resource bandwidth and availability
- Budget

Next, IT organizations need to select the most appropriate approach given the project scope. Flexibility, reusability, and total lifetime cost are the key elements to bear in mind.

Traditionally, the enterprise has been forced to choose between developing custom code or relying on an application vendor’s proprietary data migration architecture. Unfortunately, both approaches present severe challenges to cost-effective, timely, and successful implementations.

\(^2\) AMR Research, *ERP Upgrades are Big Business*, March 1, 2002

Data Migration with Custom Code

Custom code can initially offer some degree of flexibility. However, as the number and complexity of integration points increase, organizations are quick to realize all the hidden costs associated with a custom-coded approach. These costs include the creation and administration of up to 1,000 interface tables, manual data quality management processes, and the inability to easily adapt to changes. Another major weakness of a custom-coded approach is the lack of reuse—most, if not all, of the interfaces created during the original implementation cannot be reused.

Data Migration with Application Vendor’s Solution

Alternatively, while an application vendor’s proprietary data migration solution can provide some cost benefits, this approach, by definition, sacrifices flexibility in favor of limited functionality within a closed, “black box” architecture. It also carries significant risk because vendors rightly focus on delivering new application functionality versus supporting data migration utilities, which can lag behind in terms of availability, reliability, and depth. This can impact deployment schedules and resource allocation commitments. For these reasons, industry analyst Gartner Inc. has predicted that, by 2005, less than 40 percent of large market ERP II users (ERP II is defined as those enterprises with annual revenues more than $800 million) will be using their main ERP vendor’s integration layers as their primary integration platform.4

The Informatica Approach

Informatica offers a third approach to data migration. The Informatica solution is a proven, end-to-end data integration solution capable of supporting any data migration initiative, regardless of size or complexity. These initiatives span new application implementations, upgrades, and ERP instance consolidation projects.

![Figure 1](image-url)

Increasingly, organizations are turning to market-leading ETL tools such as Informatica to solve a wide variety of integration challenges, including data migration.

4 Gartner Inc., ERP Improvement and ERP II Deployment—Maximizing Return on Investment, September 18, 2002
THE INDUSTRY STANDARD
Informatica delivers the industry-leading solution for data migration by enabling organizations to:

- Accelerate development with secure, native access to legacy data, built-in productivity tools, and an easy-to-use development environment
- Reduce project risk and costs by leveraging a mature, robust and cost-effective data migration platform
- Leverage IT infrastructure and investments in existing data and business logic, along with hardware, software, storage, and security systems, through an open, standards-based architecture
- Increase operational efficiency with a metadata-centric architecture that promotes reuse at every level
- Help ensure successful user adoption and acceptance through built-in data quality features.

The Informatica Data Integration Platform: The Complete Solution

Metadata-centric Architecture Enhances the Value of Information

By abstracting definitions, procedures, and integration logic during the integration process, metadata simplifies and accelerates data migration efforts in several ways. Comprehensive metadata:

- Promotes reuse at every level
- Simplifies data validation
- Optimizes network and processor loads
- Documents all migration processes
- Captures and stores integration logic

Reusability Accelerates Development

Metadata is at the core of the Informatica product architecture. Informatica leverages metadata across the entire data migration lifecycle by making all integration routines, including source and target data definitions, integration logic, and transformation objects and procedures reusable. By reusing vital integration components, enterprises are able to leverage fully-validated, debugged, and tested data integration processes. This helps drastically reduce data migration project timelines and integration costs. Through such metadata reuse strategies, Informatica customers have been able to greatly accelerate deployment by leveraging up to 60 percent of development work.

Metadata also accelerates the application upgrade process. By storing and automatically documenting all integration logic and integration procedures through metadata, enterprises are able to quickly identify, map, and compare their new data conversion requirements to support application upgrades to existing integration routines.
Impact Analysis and Data Lineage Provide Visibility

Metadata lineage and impact analysis enable both technical and business users to quickly decipher specific source system tables to determine where data originated, when data was moved, and all the transformations applied to the data through an easy-to-use, Web-based analytical dashboard. This helps quickly validate the accuracy of data and the impact of changed data on business processes. This also provides business and technical users with an always-on view into the meaning and content of data supporting their new application.

Operational Metadata Maximizes Efficiency

Because most migrations occur over designated batch windows, it is also imperative for developers to optimize and test operational efficiencies. Through operational metadata, Informatica empowers developers with full visibility into the integration process through a variety of canned and customizable dashboards—a key requirement due to the volume and complexity of data migration projects. Operational metadata can be used to report the numbers of rows loaded into target systems, failed rows, table counts, and load times. This allows developers to maximize data migration efficiency throughout the development environment.

Flexibility Enables Migrations in Batch or Real Time

There are a variety of migration strategies employed by enterprises today. The “big bang” approach takes the production environment offline for a brief cutover period so that data can be migrated into the new application. Alternatively, enterprises may choose to adopt a “staged” approach, where new or upgraded application modules are brought online one or two at a time. While both approaches are mostly batch-driven, after migrations, real-time functionality is often required to synchronize data between systems for a period of up to a full year.

The Informatica Data Integration Platform™ is uniquely equipped to meet all batch or real-time data migration requirements, offering a wide variety of features to help ensure maximum throughput. These include pipeline and data-smart parallelism, along with automated and user-defined partitioning. To support real-time data migration and synchronization requirements, Informatica leverages its Zero Latency (ZL) engine, which provides always-on, always-listening sessions.

The result: increased flexibility to allow enterprises to adopt any data migration approach—big bang, staged, batch or real-time—rather than be limited by a particular product, or require multiple products.
Built-in Data Quality Helps Ensure User Adoption and Acceptance

Data quality is one of the most overlooked and critical issues plaguing data migration projects. Indeed, industry research has reported that up to 50 percent of application implementations fail as a result of poor data quality. Poor data quality leads to lack of user confidence in data and systems, which leads to poor user adoption.

As the enterprise migrates data into new applications, it needs to guarantee that only accurate, consistent, and complete data is loaded. Incomplete, inaccurate, missing, and duplicate data fields must be identified and normalized prior to loading. Organizations that don’t address data quality issues as part of the data migration process spend a great deal of time and cost correcting data quality issues after the migration through “band-aided” and productivity-draining efforts.

Informatica comes with built-in data quality features to solve the “dirty data” problem. Informatica offers a comprehensive library of data transformations, and a wide variety of exception reporting and validation features. For highly complex data quality requirements, Informatica partners with best-of-breed data quality vendors. Tight integration with market-leading solutions seamlessly delivers state-of-the-art data cleansing routines as pre-session, in-stream or post-session processes.

Iterative Methodology Accelerates Development

Data migration projects are extremely complex, highly unpredictable, resource intensive, and can take years to complete. Project specifications shift constantly due to unanticipated complexity, changing business requirements, and unfamiliarity with underlying source data models and systems. By adopting an iterative approach, developers are able to identify hidden requirements, compare heterogeneous source and target schemas, test performance, and constantly fine tune integration processes. The iterative development approach can drastically reduce data migration testing and mapping time.

Using Informatica’s object-based architecture, developers are able to visually map and compare source and target data definitions; quickly develop, test, and validate data migration processes from source to target systems; and batch migration components into logically constructed groups and subgroups. This helps ensure that all mappings are fully tested and validated prior to the cutover date. Moreover, all integration processes can be automatically triggered through a built-in job scheduler.
The highly iterative Informatica development environment helps the enterprise drastically reduce time to completion for data migration projects, often by as much as 12 months.

Figure 2
The easy-to-use Informatica graphical user interface (GUI) enables developers to drastically reduce data migration testing and mapping time.

Built-in Transformations and Ease of Use Shorten Lead Times

Transformations are a fundamental requirement for most data migration efforts. When moving from source to target applications, database systems, or file formats, data must be normalized and made consistent with the target environment.

For example, every business can have a unique concept of the data that constitutes a “customer” entity. Components such as birth date, address, and gender may be stored in a variety of legacy systems, with unique data formats. However, the new or upgraded application may require that the complete customer entity reside in a single table. Complex data transformations are required to reconcile the data and semantics between systems. Such transformations are also used to cleanse, parse, aggregate, summarize, merge/purge, and incorporate data from lookup tables.
The Informatica Data Integration Platform offers a comprehensive library of rich transformation objects, which reduce or entirely eliminate the need to write custom code, and perform the most complex transformation requirements. These objects handle a broad range of data management functions, allowing developers to easily parse, sort, filter, perform calculations and date checking, convert currencies, and concatenate fields.

Informatica also offers a wide variety of productivity tools to shorten development cycles, including: a drag-and-drop GUI interface, workflows, visual debugger, automated documentation, and developer collaboration tools.

Open, Flexible Architecture Leverages Existing Investments

Building a flexible and cost-effective solution is key to a successful data migration initiative. The Informatica Data Integration Platform is an open, standards-based platform that runs on UNIX, NT, and LINUX, and enables enterprises to leverage all IT infrastructure investments. This includes investments in software, storage, hardware, resources, security, and database platforms. In addition, Informatica can run in a shared server environment—eliminating the need for additional hardware.

Unlike proprietary and vendor-specific solutions, the open and robust Informatica Data Integration Platform supports native access into virtually every data source or target, including XML, EAI, legacy, relational, ERP, CRM, and SCM applications.

“Using Informatica, Wells Real Estate Fund was able to integrate 19 years’ worth of data from ten legacy databases into its Siebel CRM system. The company also synchronized data between Siebel and its custom investor accounting application. Through this large-scale CRM and investor accounting data migration and synchronization effort, Wells improved data quality, dramatically increased performance and achieved more integrated and complete customer views. Using Informatica’s data-quality features, Wells solved over 97 percent of existing data quality issues by ensuring its systems are updated with complete and accurate information. Additionally, through Informatica’s metadata-driven engine, the company has been able to reuse over 40 percent of its integration processes to accelerate development, while also leveraging metadata for impact analysis.”

— Barry Cohen
Vice President of Applications Management
Wells Real Estate Fund
Summary

Since the data migration phase of a new application implementation or instance consolidation often underpins a $10-20 million initiative, enterprises should look to proven migration solutions to ensure projects stay on budget and are successful. This is especially important in data migration projects because of their traditional high rate of failure.

Informatica is the industry-leading data integration solution, with more than 1,700 customers. With built-in quality, metadata management, flexible performance, and rich transformation capabilities, Informatica is uniquely positioned to solve all data migration requirements through a unified product.

KEY INFORMATICA BENEFITS

- Accelerated development
- Built-in data quality
- Flexible performance
- Comprehensive metadata management
- Robust transformation library
- Iterative development environment
- Proven migration solution