

# 10 Things You Need to Know Before Modernizing Your Applications

Follow these best practices to ensure your data stays clean, safe, and connected

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 March 2014

## Table of Contents

<b>Introduction</b> . . . . .	<b>2</b>
<b>Best Practice No. 1: Understand process and inter-application dependencies.</b> . . . . .	<b>3</b>
<b>Best Practice No. 2: Clean and standardize data.</b> . . . . .	<b>4</b>
<b>Best Practice No. 3: During design of your new application, build in transparent data access and integration points to avoid data lockdown.</b> . . . . .	<b>5</b>
<b>Best Practice No. 4: Sync with existing apps and data warehouses using the most current data</b> . . . . .	<b>5</b>
<b>Best Practice No. 5: Classify data and define data retention and privacy requirements.</b> . . . . .	<b>6</b>
<b>Best Practice No. 6: Leverage an automated test data provisioning process</b> . . . . .	<b>7</b>
<b>Best Practice No. 7: Leverage MDM to provide a consistent single source of enterprise master data for current and future transformation initiatives.</b> . . . . .	<b>8</b>
<b>Best Practice No. 8: Know your data integration team and what tools they use</b> . . . . .	<b>8</b>
<b>Best Practice No. 9: Build legacy application shutdown into the plan for when you go live</b> . . . . .	<b>9</b>
<b>Best Practice No. 10: Smart partitioning for application performance</b> . . . . .	<b>10</b>
<b>Why Informatica?</b> . . . . .	<b>11</b>

## Introduction

### GOLDEN RULES

What does it mean to keep your data clean, safe, and connected?

**Clean:** Data contains no errors or omissions

**Safe:** Data is shared only with authorized persons (and applications)

**Connected:** Data flows freely among the applications that need it

*Transformation* is the business buzzword of the day. It's overused. But in the case of what's happening to enterprise applications in businesses around the world, it's most definitely appropriate.

Applications are being transformed at rates never seen before. Seventy-one percent of enterprises said that IT is deploying more, and more complex, applications than it was a year ago.<sup>1</sup> Businesses are integrating on-premises applications with cloud applications such as Salesforce (28 percent of enterprises did so in 2013) or moving them to the cloud (40 percent of enterprises).<sup>2</sup> Applications are being consolidated due to a once-again robust M&A market, where aggregate M&A deal value for 2013 was \$958 billion, the second-highest amount spent since the 2008 financial crisis.<sup>3</sup>

Then you have the transformation necessary when moving from one application to another, such as from SAP to Oracle, or vice versa. Or to upgrade to the latest version of an ERP application. Or to implement master data management (MDM).

Integration is always a part of application transformation: making sure new systems talk to existing systems and vice versa. Gartner predicts organizations will spend one-third more on application integration in 2016 than they did in 2013. What's more, by 2018, more than half the cost of implementing new large systems will be spent on integration.<sup>4</sup>

Whatever your reasons behind application transformation, a critical part of any initiative is migrating data from an old environment (or environments) into a new one. For large businesses, this is not cheap. On average, such a migration costs \$875,000.<sup>5</sup> If the migration is part of a larger project—as it typically is—the average cost is \$2.8 million. The average cost of a project overrun is \$268,000 or approximately one third of a data migration project.<sup>6</sup> And an unfortunate reality is that such projects frequently experience significant cost overruns.

To avoid such overruns and other potential potholes on the road to a successful application transformation, here are 10 best practices you'll be glad you knew about before embarking on your journey.

<sup>1</sup> Cisco Global Cloud Networking Survey 2012

<sup>2</sup> Data Migration Customer Survey, Bloor Research, February 2014

<sup>3</sup> FactSet:FLASHWIRE US MONTHLY, January 2014

<sup>4</sup> "Gartner Predictions 2013 for Application Integration: My Take," by Dinesh Chandrasekhar, November, 2012

<sup>5</sup> Data Migration Customer Survey, Bloor Research, February 2014

<sup>6</sup> Ibid

## Best Practice No. 1: Understand process and inter-application dependencies

Data isn't static. It flows through your organization, through multiple applications and many, many business processes. If you don't know your data flows and dependencies, you're heading for failure during an application transformation. Even when individual application owners are masters of their own domains, if they don't understand data flows to and from other applications, they won't be able to untangle the downstream impact of any changes they make.

This is especially critical during M&As. All too often, a company won't understand the processes that exist within an M&A target and how heavily the target company depends on the data flow of such processes to function. Say an enterprise has established important workflows between its on-premises marketing automation and CRM systems and is decommissioning them to move to a new cloud application. Retiring them separately or at the wrong time may lead to all-important sales leads falling between the cracks.

It's critical, therefore, to trace data through the enterprise and document the flow, typically in asset spreadsheets. Then you can deploy metadata management tools for change control and Visio to provide visual mappings of the data flows among systems to ensure data integrity within a business context. You may even want to hire a SWAT team of data flow experts—yes, they exist, and are typically called data architects—to do this for you. Whatever method you choose, mapping data flows will be your single most well-spent use of time and money. You'll even find "overflow" value in documenting data flow—in addition to enabling application transformation, for example, understanding data flows can aid in analytics.

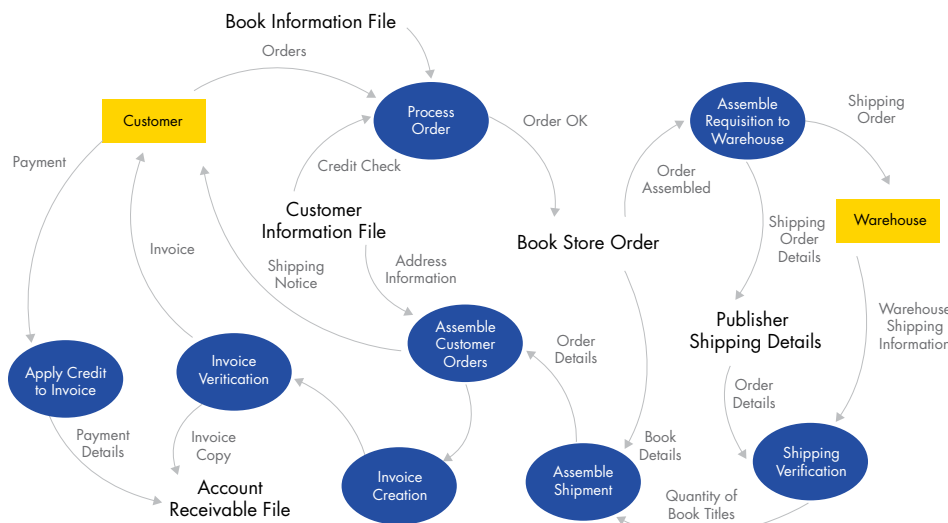
"If application integration doesn't become a true area of expertise, companies will find themselves at a serious competitive disadvantage within the next few years."

—Benoit J. Lheureux, research vice president, Gartner<sup>7</sup>

"Very few companies have thoroughly mapped their data flows and process dependencies. That's too bad, because it's relatively easy to do, and it delivers tremendous value. Without understanding how data moves through different applications and replication points, companies will suffer from gaps in business knowledge. At the end of the day, companies need to turn data into insight and insight into action, so it's pretty critical that you know where your data comes from, and where it's going."

– Tomas Kornegay, director of analytics for digital sport at Nike

Data Flow through Accounts Receivable



A data flow is typically documented in an asset spreadsheet and then rendered graphically. This ensures data integrity within a business context.

<sup>7</sup> Speech at Gartner's Application Architecture, Development & Integration (AADI) Summit in Las Vegas, Nov. 2012

## Best Practice No. 2: Clean and standardize data

Now that you know how your data flows, you need to make sure data is clean and standardized across applications. The old adage, “garbage in, garbage out” is still as relevant in the mobile, always-on, always-connected world as it was in mainframe days.

### DIFFERENT TYPES OF DATA TOOLS

**Data profiling tools:** These identify relationships among data sources across applications and discover and monitor poor quality data in those data sources

**Data quality tools:** These cleanse and de-duplicate existing data while preventing entry of new dirty data into the transformed application

**Data integration tools:** These transfer data among different applications, transforming data formatting when necessary

Data quality issues are exceedingly common. Customer records are often missing addresses, for example, or have incorrect phone numbers or ZIP codes embedded in them. If you’re not careful, this “dirty” data will be in your new system and propagated throughout the enterprise—and business executives will make decisions based on bad information. It’s therefore essential to clean and standardize data before you attempt to transform your application.

When it comes to standardization, you need to understand where the meaning of data can differ from application to application. What’s your data “truth,” and how do you get agreement among applications regarding that truth?

One way to achieve this is to employ data profiling and data quality tools. Bloor Research, in a multiyear study of data migrations that began in 2007, observed that when companies used automation to cleanse and standardize data, the success rate of migrations skyrocketed. For example, in 2007, only 10 percent of enterprises used data profiling tools (see sidebar for definitions). In 2011, 72 percent of organizations did.<sup>8</sup> Projects brought in on time and within budget rose from 16 percent to 62 percent. During the same period, use of data cleansing tools rose from 11 percent to 75 percent, which presumably contributed to the dramatic success of the projects.<sup>9</sup>

You can also take advantage of data migration tools that provide “before” and “after” comparisons of data, allowing you to verify that the migrated data is clean and accurate.

Additionally, business glossaries can help by providing definitions and descriptions of data. Not only will this help with data migrations between applications—you can then create data quality rules that automation will enforce during transformation—but it will also help business and IT work better together, because they will be using a common terminology. For example, agreeing on what constitutes a “customer” from both business and IT perspectives is extraordinarily important.

And make sure to incorporate ongoing data quality rule validations and management into your new application. If you don’t, your data quality will erode 1 to 1.5 percent a month, according to Bloor Research.<sup>10</sup>

<sup>8</sup> Data Migration Customer Survey, Bloor Research, February 2014

<sup>9</sup> Ibid

<sup>10</sup> “Data Migration,” Bloor Research, May 2011, Philip Howard

## Best Practice No. 3: During design of your new application, build in transparent data access and integration points to avoid data lockdown

Data lockdown occurs when data is isolated in an application silo and can't be shared easily with other applications. Given the increased emphasis on improving operational efficiencies, analytics are moving into operational applications. For front-line managers to get the kind of visibility that enables them to react in a more timely fashion to new trends, they need real-time access to data. In legacy applications or platforms where data integration was not considered upfront, providing access to data required customizations that involved lengthy development cycles. Today, such development cycles are considered unacceptable. Real-time access is the new norm.

How to achieve this? Build open interfaces and integration points into your new application so data can easily move back to existing applications—even legacy ones. And build those integration points with the plan to integrate even more data in the future from applications you may not even know about yet.

Still, integrating all these data points manually is challenging. Happily, tools are available that make such data integrations easy without the need to write customizations. As you replace legacy applications, or integrate legacy applications with modern ones, look for ways to leverage a standard so that customizations can be avoided at all costs.

## Best Practice No. 4: Sync with existing apps and data warehouses using the most current data

When it's time to populate a new application with data, you want to make sure you are migrating the most current data from the correct "systems of record" within your organization. In many cases, such systems are dependent on others (see Best Practice No. 1). Even transactional systems, which are theoretically up to date at all times, can depend on data from outside applications. For that reason, you need to synchronize all relevant data from existing applications and data warehouses before migrating it to your new, transformed, application.

For example, if one company has acquired another, it can end up with hundreds of older, redundant systems that it wants to transform to a single new one. Step 1 during data migration is to ensure that it has correctly identified the systems of record for each data element. Step 2 is to ensure that the data in those systems of records is current by identifying the chain of dependence on other applications.

Then, when you develop test beds for the new application, make sure you have an automated way to checkpoint and refresh test data. When you go live, you shouldn't be surprised by gaps in data, as that could result in a skip in a business process and the loss of important transactions. Define a checkpoint in the plan so this step doesn't get overlooked. This will avoid duplicate or missing data sets.

"The first time we started doing some data migration loads, and we saw the amount of transaction failures as a result of missing data, was very eye-opening... As a result of that, we implemented a data quality project that had dashboards to measure data readiness for a big SAP environment, and it really was eye-opening for the team to see the business impact of the data, both on the project, and on other functions downstream."

– Barbara Latulippe, senior director enterprise information quality & MDM, EMC

## Best Practice No. 5: Classify data and define data retention and privacy requirements

In many application transformation projects—especially those driven by M&As—you will be decommissioning now-redundant applications that support a common business process. Yet that creates challenges because business and regulatory compliance reasons may require that the data currently stored in those applications may need to be archived and stored for a longer length of time. Some industries—particularly financial services and healthcare—have very stringent standards for data retention requirements. Indeed, a recent Informatica survey found that 60 percent of all application migration projects require some degree of archival.

This is a common challenge. For enterprises with more than 100 databases, 61 percent anticipated decommissioning business applications over the next 12 to 18 months.<sup>11</sup>

In the past, enterprises typically used data's age as the way to determine whether to retain it. Older data was either deleted or transferred to tapes or some cheaper—and less accessible—means of storage without regard for the value it might still offer the business. Today, this is not an option, as enterprises are under scrutiny to both increase data security and retain data to meet compliance regulations. Planning ahead for a cost-effective and efficient data classification strategy is therefore an imperative during an application transformation.

Creating a classification strategy typically involves collecting retention and business-user access requirements to ensure that critical data is still accessible to users once archived. Enterprises that are primarily concerned about compliance should create a data classification schema that maps out which fields in a database are most sensitive. You should also identify any dependencies the data may have with other stored information (see Best Practice No. 1).

Additionally, most application transformation projects need to mask data for security and compliance reasons. Doing this manually via hand-coding is rarely an option for time and cost reasons, and because of the potential for errors. And although most data quality and integration solutions offer limited masking capabilities, you should choose an automated data migration platform that incorporates best-of-breed data masking capabilities. Make sure that whatever solution you choose offers a full audit trail so you can prove that you have masked the data you need to have masked.

<sup>11</sup> "Application Retirement Trends," Enterprise Strategy Group, October 2011



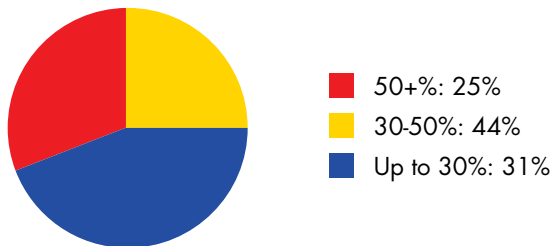
## Best Practice No. 6: Leverage an automated test data provisioning process

Testing is an essential part of the application transformation process. It's also the most time-consuming—and therefore costly part. Gartner estimates that development and testing takes the most amount of time when developing a new application, consuming 24 percent on average of the entire application development lifecycle.<sup>12</sup> And a recent Informatica survey found that the tasks involved in managing test data—such as defining which data to use in a test case, creating or copying test data, and securing sensitive data—takes more than half of developers' and the QA team's testing time.<sup>13</sup>

Manually creating appropriate test data usually adds up to significant dollars, representing a great deal of inefficiency and waste in application transformation budgets. If your organization has multiple application development projects underway, this could mean tens of millions of dollars of testing resources that could more effectively be used elsewhere. Moreover, manually writing scripts to create test data is prone to errors. Organizations that installed Informatica's Test Data Management (TDM) saw an average of 30 percent or more improvement in testing time, according to a recent Informatica survey (see graphic).

According to users of Informatica Application ILM in answer to the question, What were the time savings from using Informatica data masking or TDM tools?

### Automated Test Data Management Saves 30% More Time



Source: Survey of 32 users of Informatica Application ILM

By using an automated solution, you eliminate manual processes in favor of self-service ones. You get significantly improved test data quality—with fewer errors or defects—at dramatically reduced costs. And you lay the groundwork for future application transformations.

Many enterprises make the mistake of simply copying production data for testing purposes. But if you've outsourced your testing overseas or your test teams reside outside the firewall or in another country, you cannot use sensitive data in nonproduction, unprotected environments. You need to ensure you are in compliance with all local data privacy and data residency laws. The European Union, for example, has strict data privacy rules that you must adhere to. In some cases, test data will simply not be available, and you will need the capability to synthetically generate test data.

<sup>12</sup> Gartner IT Key Metrics Data, December 2011

<sup>13</sup> TechValidate. TVID: 988-ED5-CCD

MASTER DATA  
MANAGEMENT:  
A DEFINITION

A technology-based discipline in which business and IT work together to ensure the uniformity, accuracy, stewardship, semantic consistency, and accountability of an enterprise's official shared data assets.

## Best Practice No. 7: Leverage MDM to provide a consistent single source of enterprise master data for current and future transformation initiatives

Migrating data is never a one-time event. Yet all too often during application transformations, data migration is treated as a one-off task. This can lead to higher costs, delayed deployments, longer time-to-value, and the risk of not meeting business objectives.

By using master data management (MDM), enterprises can minimize risk and speed data migration. You should choose an MDM solution that cleanses, standardizes, and enriches dirty data; removes duplicates and creates one version of the truth; and centrally manages data cleansing rules—all during the pre-migration stage. During migration, the ideal MDM solution will simplify the migration architecture by avoiding rigid models and structures that don't allow you to adapt as your business changes. It also automates management of very large volumes of data. Finally, during post-migration, a good MDM solution will maintain consistent data across systems.

The most critical aspect when choosing an MDM solution is that it is application-agnostic, repeatable, and prepares you for your next migration project. You will consistently be able to deliver quality data results, on time, and on budget. You'll also reduce the cost of every incremental migration. And you'll be able to scale as your business grows.

## Best Practice No. 8: Know your data integration team and what tools they use

This is an important but frequently overlooked best practice. Managers in charge of application transformation projects must get to know their data integration teams. They need to know the skill sets of the members of those teams and the tools those individuals plan to use for the integration. After all, just because someone happens to be master of a particular application doesn't mean he or she will be able to transform the data into a new application in a way that ensures it remains clean, safe, and connected.

Some businesses search for data scientists to do this part of the application transformation job. Unfortunately, true data scientists are in short supply—a situation that is only expected to worsen in coming years as Big Data comes into its own.<sup>14</sup> Data scientists are experts who understand the big picture of data transformation and are capable of designing the right models and algorithms to derive meaning from vast quantities of information. But do you really need a data scientist for an application transformation initiative? More businesses are finding out that the answer to that is a resounding “no.” Instead, you can effectively transform your data by having the right team and tools in place to cover all the skill sets.<sup>15</sup>

<sup>14</sup> “It Takes Teams to Solve the Data Scientist Shortage,” by Jeanne G. Harris, Nathan Shetterley, Allan E. Alter, and Krista Schnell, *Wall Street Journal*, February 14, 2014

<sup>15</sup> “How To Build A Successful Data Science Team,” by Jeff Bertolucci, December 20, 2013

So who do you need on your team? The answer is a combination of business analysts, project managers, statisticians, systems architects, and data engineers. You might even consider hiring graphic artists or designers to help with data visualization.<sup>16</sup> The good news is that you may already have all of these skill sets in-house. Your job then becomes putting them together into a cohesive team, and managing them appropriately.

Note that your data integration team most likely possesses quite a bit of data-flow knowledge and skill sets that can be leveraged during a migration—especially since the tools of data migration are typically the same as for data integration.

## Best Practice No. 9: Build legacy application shutdown into the plan for when you go live

We are all creatures of habit. Studies show that human behavior is 93 percent predictable, and that we won't change our behavior unless forced to.<sup>17</sup>

So one way to ensure the new application gets adopted quickly is by ripping off the band-aid. Don't allow redundant applications to run in parallel. Once you have transformed your application—whether through consolidation, upgrade, or retirement—turn off the old one. The reason? Users simply won't adopt the new application until the old one is no longer available. And the cost of keeping legacy systems running can be exorbitant. By turning off old systems, you free up IT dollars for innovations that can more directly drive business success.

What creates redundant legacy applications? Upgrades and consolidations are the top drivers, with 33 percent of enterprises saying that this is the reason they end up running systems in parallel.<sup>18</sup> Twenty-two percent said that M&A activities created their redundant legacy applications.<sup>19</sup>

Redundant systems cost you—a lot. Approximately 75 percent of organizations spend at least \$100,000 annually supporting legacy applications. Larger enterprises spend even more on people, software licensing, and maintenance expenses.

<sup>16</sup> "It Takes Teams to Solve the Data Scientist Shortage," by Jeanne G. Harris, Nathan Shetterley, Allan E. Alter, and Krista Schnell, Wall Street Journal, February 14, 2014

<sup>17</sup> "Human behavior is 93 percent predictable, research shows," PhysOrg.com, February 23, 2010

<sup>18</sup> "Application Retirement Trends," Enterprise Strategy Group, October 2011

<sup>19</sup> Ibid

## Best Practice No. 10: Smart partitioning for application performance

It's a fact that data volumes are growing. Data production will be 44 times greater in 2020 than it was in 2009.<sup>20</sup> And enterprises are challenged to keep application performance acceptable to users given the velocity of growth. Given that IT budgets are already strained, purchasing more—and more powerful—hardware to maintain performance isn't typically an option. Some organizations archive data based on a classification scheme to minimize the weight of data on systems. But this isn't always helpful, as the volumes of active data that remain are still too large. The result is that performance of mission-critical databases slows, and business decision-making is correspondingly delayed, often to the detriment of the enterprise.

The answer? Smart partitioning.

Smart partitioning involves physically organizing data in the database to optimize performance. When using an automated solution, partitions can be created based on any number of parameters or complex rules—whatever makes sense based on user requests. Smart partitioning also streamlines archiving of data as it becomes less relevant to the business.

With smart partitioning, you get the same performance boost as if you had deleted data from your database. You can extend the life of your investment in hardware and streamline database operations. Small wonder that Oracle recommending smart partitioning to enterprise customers of its Oracle Database.

<sup>20</sup> Big Data Universe Beginning to Explode, CSC, January 2013

## Why Informatica?

To ensure a successful outcome to an application transformation initiative, you need to remember we are in the age of data-centricity. If you don't put data first, you end up spending more time and wasting money.

Technology tools exist to streamline data management from creation through maintenance, modernization, and retirement. Informatica is here to help you leverage the best-in-class technology built on a proven platform that enables you to leverage available skills in the market.

The key components you will need in your toolbox are:

- Application connectivity
- Data profiling
- Test data management
- Data masking/de-identification
- Data quality
- Data integration
- Master data management
- Smart partitioning and data archive
- Application retirement

Informatica offers a portfolio that encompasses all these toolsets. It has supported more than 10,000 application and data migration projects and leads the industry in helping enterprises successfully transform their applications for competitive advantage. Planning for applications modernization? Let's talk.

## 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.



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

© 2014 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.