Best Practices for Data Migration and Data Management

Businesses run on data, and today that data is moving to the cloud, creating new opportunities for digital transformation and data management.
# Table of Contents

4  **CHAPTER 1: 7 BEST PRACTICES FOR MIGRATING YOUR DATA TO THE CLOUD**

13 **CHAPTER 2: REIMAGINING ETL FOR A CLOUD-CENTRIC WORLD**

19 **CHAPTER 3: DATA WRANGLING: ENSURE CONSISTENCY ACROSS APPLICATIONS AND CLOUD PLATFORMS**

26 **CHAPTER 4: GIVE YOUR BUSINESS A LIFT WITH ‘DATA GRAVITY’**

30 **CHAPTER 5: 7 BEST PRACTICES FOR BUILDING DATA QUALITY INTO INTEGRATION**

39 **CHAPTER 6: MULTIMASTERING DATA FOR APPLICATION MIGRATION**

43 **THE BIG PICTURE: DATA MANAGEMENT DONE RIGHT**
INTRODUCTION

This is a book about moving data to the cloud and then ensuring that a business makes the best possible use of that data.

It’s about the importance of putting data first – recognizing that data should drive your decisions about applications and services, not the other way around. It’s about developing strategies for synchronizing data across applications and data stores, and why one approach can deliver better results than another.

It’s also about recognizing that times have changed. The data paradigms of the past, such as Export Transform and Load (ETL), have been a bedrock of data migration and data analysis for the past several decades. But given new cloud-centric businesses and IT initiatives, they’re no longer the default approaches. Technologies such as the integration platform as a service (iPaaS) offer faster delivery times and more flexible options.

More than ever before, businesses have choices. They can mix legacy and cloud technologies, take advantage of both traditional and new data repositories, and process data wherever it’s needed: on-premises, in the cloud, or at the network edge.

To make the right choice, you need to ask the right questions – and understand the most promising possibilities.

This eBook, a joint project of Boomi, a Dell Technologies business, and Slalom, is for CIOs, enterprise architects, integration engineers and other IT stakeholders. All will find the guidelines they need to manage and make the most of their data in the cloud.

Let’s dive in.
By Oliver Asmus, data & analytics leader, Slalom

Think about cloud migration, and you probably think first about applications. You might have an on-premises application such as an enterprise resource planning (ERP) system that you want to replace with a SaaS application performing a similar function. In this scenario, once the cloud application is up and running, the project is complete.

But from our experience at Slalom guiding organizations through cloud-migration projects, we’ve learned that it can make more sense to think first about migrating data. That is, an organization should first figure why and how it’s migrating data to the cloud. Only then should it turn to the question of applications and data services.

Data is the most important asset of nearly every company today; in our digital economy, data is the “new oil.” You might want to move data to the cloud so it can be used by a variety of cloud applications and data services, and not just by your on-premises applications. But if your migration project is limited to replacing on-premises applications with their cloud equivalents, you’ll likely overlook many of your data’s new cloud-enhanced use cases.

By focusing instead on data, you can think more broadly and creatively. Focusing on data can also help you to lay the groundwork for data management in the cloud. That’s the key to successfully modernizing and transforming your digital business.

Let’s get started by considering seven best practices for migrating data to the cloud.

Chapter One
7 Best Practices for Migrating Your Data to the Cloud

When migrating to the cloud, put data first. Adopting a “data first” strategy will help you lay the groundwork for data management – controlling data and ensuring data quality – in the cloud.

By Oliver Asmus, data & analytics leader, Slalom

Think about cloud migration, and you probably think first about applications. You might have an on-premises application such as an enterprise resource planning (ERP) system that you want to replace with a SaaS application performing a similar function. In this scenario, once the cloud application is up and running, the project is complete.

But from our experience at Slalom guiding organizations through cloud-migration projects, we’ve learned that it can make more sense to think first about migrating data. That is, an organization should first figure why and how it’s migrating data to the cloud. Only then should it turn to the question of applications and data services.

Data is the most important asset of nearly every company today; in our digital economy, data is the “new oil.” You might want to move data to the cloud so it can be used by a variety of cloud applications and data services, and not just by your on-premises applications. But if your migration project is limited to replacing on-premises applications with their cloud equivalents, you’ll likely overlook many of your data’s new cloud-enhanced use cases.

By focusing instead on data, you can think more broadly and creatively. Focusing on data can also help you to lay the groundwork for data management in the cloud. That’s the key to successfully modernizing and transforming your digital business.

Let’s get started by considering seven best practices for migrating data to the cloud.

Chapter One
7 Best Practices for Migrating Your Data to the Cloud

When migrating to the cloud, put data first. Adopting a “data first” strategy will help you lay the groundwork for data management – controlling data and ensuring data quality – in the cloud.

By Oliver Asmus, data & analytics leader, Slalom

Think about cloud migration, and you probably think first about applications. You might have an on-premises application such as an enterprise resource planning (ERP) system that you want to replace with a SaaS application performing a similar function. In this scenario, once the cloud application is up and running, the project is complete.

But from our experience at Slalom guiding organizations through cloud-migration projects, we’ve learned that it can make more sense to think first about migrating data. That is, an organization should first figure why and how it’s migrating data to the cloud. Only then should it turn to the question of applications and data services.

Data is the most important asset of nearly every company today; in our digital economy, data is the “new oil.” You might want to move data to the cloud so it can be used by a variety of cloud applications and data services, and not just by your on-premises applications. But if your migration project is limited to replacing on-premises applications with their cloud equivalents, you’ll likely overlook many of your data’s new cloud-enhanced use cases.

By focusing instead on data, you can think more broadly and creatively. Focusing on data can also help you to lay the groundwork for data management in the cloud. That’s the key to successfully modernizing and transforming your digital business.

Let’s get started by considering seven best practices for migrating data to the cloud.
PRACTICE ONE

Define your goals

To begin, identify the outcome you expect to achieve by migrating your data to the cloud. Is that outcome better performance? Improved scalability? The opportunity to take advantage of new technologies available only through cloud services?

If you’re not sure which new technologies are available in the cloud, explore that topic beforehand. Do some research, looking into how you can make the most of either the data your organization already has or that it routinely collects.

Next, document the goals driving your migration. It’s fine if your plan includes some sweeping, long-term initiatives. But be sure to also include short-term goals that are readily achievable. That way, your project can quickly demonstrate value and build much-needed momentum.

If you set a sweeping goal—such as replicating a 30-year-old on-premises infrastructure in the cloud within a year—you’re probably going to be disappointed. Instead, try to set goals that are not only important, but also achievable. Then let them serve as the first steps for additional migration work in the future.

The cloud offers tons of benefits. But don’t underestimate the amount of due diligence it takes to ensure a successful migration.

By focusing on data, you can think more broadly and creatively.
Don’t assume that when moving to the cloud, you should simply “lift and shift.” If, for example, you’re running a data warehouse on-premises in Oracle, and you’re interested in taking advantage of the cost/performance benefits of the cloud, don’t assume you should simply move your data to Amazon Redshift, the data warehouse service available through Amazon Web Services. To fully leverage the cloud, you might do better by combining the data warehouse with other data services.

Determine which business results you’re really trying to achieve. Then work with an enterprise architect to design the cloud services that will deliver those results. Migrate to the architecture that will meet your future needs, too. It might end up being a mix of applications and services, even if your on-premises implementation comprises only a single application.
PRACTICE THREE
Work with a cross-disciplinary team right from the start

As you plan your migration to the cloud, you’ll need the support of a multidisciplinary team. Don’t talk only with data scientists, application developers and cloud architects. Also include your security and networking teams. That way, you’ll ensure that your cloud migration makes sense for the organization’s broadest data-management needs.

Also include leaders from business units that will be affected by the migration. You may want to enlist members of the executive team who can champion your project, too.

Recognize that the members of your team may have competing priorities. Try to get their buy-in early on.
Determine whether your data would benefit from a multicloud strategy

In most data-migration projects, a company moves data to applications and storage hosted by a single public-cloud service, such as Amazon Web Services (AWS), Google Cloud Platform or Microsoft Azure. But sometimes it makes sense to move data to services spread across multiple vendors. That’s what’s known as a multicloud strategy.

If your data could benefit from a multicloud strategy, it’s good to figure that out upfront and then plan accordingly.

Why adopt a multicloud strategy?

Good reasons include:

**Unique services:** One cloud vendor might offer a special application or service that’s worth using, even if other vendors do a better job handling other applications or services.

**Better price/performance:** One application or service offered by a vendor might be so much more affordable, or deliver so much better performance, that it’s worth adopting, even if it’s not completely unique. Meanwhile, you can let other applications and services be handled by another vendor.

**Industry experience:** Some vendors have lots of experience meeting the exacting requirements of a special industry or market. For example, Microsoft has a strong record offering cloud services that meet the data-privacy and security requirements of the government. If a vendor has expertise in your organization’s area, then you might entrust it with certain data and applications. You might also choose another vendor for less specialized services.

**Regulatory compliance:** Is the data you’re migrating to the cloud subject to a regulation, such as the European Union’s General Data Protection Regulation (GDPR), that calls for “data localization,” the storage of data within a particular nation or region? If so, then you might want to set up a cloud architecture with independent instances in multiple regions. That might mean working with one vendor for Europe, another for North America, and so on.
PRACTICE FIVE

Identify data discrepancies – and clean them up – before you migrate

Before you migrate data to the cloud, you’ll want to make sure the data is accurate, complete and consistent. That means identifying any data discrepancies and then cleaning them up, and well before the migration.

Some of our clients have run into the problem of "shoehorning" data, which involves the use of a single data field to hold different types of data. For example, a CRM record might have a field that has no governance behind it. As a result, one person might use it for phone numbers, while another person uses it for street addresses, and yet another uses it for last names. Shoehorning typically occurs when there's only one unlocked field for data in a table. Since there's only one choice, each user enters whatever data he or she finds most important.

But if you're going to move that data to the cloud, you'll need to clean it up first. Phone numbers belong in phone-number fields, last names in last-name fields, and so on. Also, some of that data might be subject to data-privacy rules and regulations. If so, you definitely need to know where that data is and how it's being stored.

Before migrating any data to the cloud, audit the data, identify any inconsistencies and then resolve them. Otherwise, you risk migrating data discrepancies to the cloud and perpetuating the problem there. Also be sure to identify any data fields that need to be created or changed in your cloud data-storage system. Agree on rules for data entry and data governance. Finally, note any data fields that you'll need to track for ongoing regulatory compliance.
Seek a platform that minimizes “throwaway” work

With any substantial data-migration project, some amount of coding will end up being “throwaway” code. For example, some software might perform a one-time, never-to-be-repeated transformation. Or you might need to move data to a temporary system or location before completing the migration. Especially in large enterprises with lots of legacy systems, it’s almost inevitable that you’ll need some of this one-off coding.

But just because throwaway code is necessary doesn’t mean you need to invest a lot of time in it. Instead, write just enough code to get the job done. Then move on.

One advantage of a low-code platform for integration and migration — an integration platform as a service (iPaaS) — is that producing one-off integrations and transformations becomes quick and easy. For example, if you need to extend integrations or transformations with custom Java code, you can, and with ease.

A low-code iPaaS lets you quickly take care of your “throwaway” code. It also let your IT talent get back to focusing on your more strategic projects.

If you do need to create temporary code, you can do so quickly and easily in iPaaS. When that code is needed, its usefulness might be great. But when that time is past, the loss is minimal, because you won’t have invested a lot of time and effort in something that turned out to be temporary. Put simply, low-code development leads to both lower costs and far greater production efficiencies.
Take advantage of both integration and data management with a modern iPaaS

Some companies have become wary of the very idea of data management. That’s because they assume it will require a multimillion-dollar investment over multiple years. But modern technologies such as Boomi’s iPaaS make it possible to establish a golden record and carry out ongoing data governance far more efficiently.

One reason a unified approach to integration saves time and money is that both your integration processes and data-management system sit on the same cloud platform. This eliminates the need to send data from one cloud application down to an on-premises data-management system, then back up to another cloud application. Instead, a unified platform such as Boomi keeps everything in the cloud, using its hub-and-spoke structure to directly connect applications.

With cloud-based data governance, you won’t need to download drivers or set up a complex on-premises data-management system. Instead, you can get things going quickly and easily in the cloud – which, not incidentally, is also where you’ve migrated your data.

Also, because the iPaaS manages the integrations, you can build data management and data cleansing right into your integration process. That way, all data moving from one application to another is both accurate and complete.

In addition, the low-code development interface available in an iPaaS lets you implement data-governance policies quickly and effectively. You won’t have to embark on that multiyear, multimillion-dollar project. Instead, you’ll start seeing benefits in just months, even weeks.
We’re living in a golden era of analytics. Many companies now appreciate how the power of big data and artificial intelligence can transform their operations and help to deliver better business results. Because those analytics and operations depend on data, it makes sense to focus your strategies on data first.

By following the best practices we’ve recommended above, you’ll migrate data to the cloud quickly and effectively. You’ll also be ready with complete, accurate and consistent data to take advantage of the new cloud applications and services emerging practically every day.
ETL, a traditional approach for data integration and analysis, is too cumbersome for today’s fast-moving markets and digital enterprises. Fortunately, there’s a better way: modern integration.

By Oliver Asmus, data & analytics leader, Slalom

ETL, short for Extract, Transform and Load, is a traditional approach for moving and transforming data for analysis. It’s especially common in large organizations in regulated industries.

But for today’s modern, digital enterprise companies, ETL is often too cumbersome. Organizations want their data right away. They see no reason to invest time and money in ETL when a modern, cloud-native integration platform can deliver the data they need far faster and more efficiently.

In this chapter, we’ll review the benefits and drawbacks of ETL. We’ll also explore how an iPaaS (integration Platform as a Service) can easily overcome ETL’s shortcomings and give this integration method a turbocharged update.
ETL: Slow and steady is sometimes okay

ETL is a way of extracting data from a source system, manipulating and transforming the data as it’s being moved among systems, and then loading it into a data warehouse or other data repository.

ETL is a tried-and-proven model for moving data, particularly for projects that require a data warehouse for making data available for analysis or enriching data sets in other applications. And ETL is time-tested; it’s been around for more than 30 years. Virtually all enterprise architects learn about ETL when they begin their careers in data integration and management.

Even today, ETL is a fine approach for certain types of data integration and analysis projects. For example, a financial-services organization could use an ETL tool to extract data from various core systems, transform that data, and then load it into a data warehouse for analysis and reporting.

This example presents some of the key characteristics of a typical ETL use case:

**A repeatable, predictable process:** The process of extracting, transforming and loading data into a data warehouse is well-defined, stable and repeatable. The financial institution processes the same types of transactions day in and day out. It knows which applications are storing the data; it also knows what sort of analysis is needed from the warehouse in the end. The overall process is fixed, and the outcome is predictable.

**Focused on compliance:** The process is run in the context of finance, a highly regulated industry. In finance, a process’ stability and predictability are considered benefits. Regulators want the institution to handle its data carefully and generate the required reports. ETL’s predictable processes can help an organization meet these regulatory requirements.

**Slow is fine:** Speed isn’t critical to building this workflow or running it on a regular schedule. If an ETL integration takes three months to build, that’s okay, because once it’s built, it’s going to be used with few or no changes for several years. The results don’t need to be rushed, either. In ETL uses cases, slow and steady wins the race.

In this type of scenario, ETL processes do fine. Large organizations can be expected to continue using ETL for these types of data projects into the foreseeable future.

In many industries, we’re seeing young, cloud-centric companies skip ETL altogether and go straight to iPaaS. These companies are used to moving quickly, and iPaaS suits their needs for moving data and getting the answers they need on time.
Fast, fluid data integration with iPaaS

Now let’s consider examples from two other industries: consumer packaged goods and retail. Compared with financial services, these markets move much more quickly.

For example, in the fashion market, a manufacturer might design, make and launch a new clothing line in just a matter of a few weeks. So they have to be nimble.

In these fast-paced markets, ETL’s slow and steady integration simply won’t do. Developers in these markets will need to integrate new sources, load data into platforms for business intelligence and trend analysis, and get results back to business decision-makers, all as quickly as possible.

This use case presents a different set of characteristics:

**Processes and analytics are in flux:** Processes change constantly. Suppliers, products, partners and campaigns change from week to week, month to month, and season to season. With these changes, the questions asked by business leaders – and their answers – will change, too.

**Compliance isn’t front and center:** Repeatability and compliance aren’t as important to these industries as they are to financial services. Many of the analytical reports being produced are not required for regulatory compliance. Instead, they’re intended to support critical business decisions.

**Speed is of the essence:** Retailers, consumer-products companies and most other digital enterprises need answers as quickly as possible. Their developers don’t have weeks and months to hand-code integrations, but rather, just a day, or even an afternoon, to move the data where it needs to go.
Why iPaaS is replacing ETL in fast-moving, cloud-based markets

For companies in these fast-moving markets, iPaaS really shines. These organizations can use the Boomi Platform for integration, data governance, API management and workflow automation, rapidly tying together their applications and data.

Boomi’s low-code interface and ready-to-use connectors bring unprecedented speed to unite and transform your data. You can:

**Complete projects in days or hours** that with traditional hand-coding techniques and legacy middleware might have taken a month.

**Load data from any source quickly** into a business-intelligence application such as Qlik or a cloud data warehouse such as Snowflake.

**Export analytical results to any application.**
Modern integration helps digital enterprises thrive

At Slalom, we’re hyper-focused on partnering with our customers. We see many companies across a wide range industries — from education to telecommunications to retail — sharpening their customer focus, too. But this increased attention also requires raw data and data analytics to answer questions that include: How well are our campaigns working? Do customers like our new features? Are customers actually using the products, and if so, how? In which regions and demographics are we doing best?

To answer these kinds of questions, a business needs to be agile, curious and open-minded. An answer might require you to examine a data set six different ways. You might need to connect data from Salesforce, a retail store’s real-time analytics application, a mobile service and survey data. You might need to run different types of analytic algorithms, then follow up with more questions to find even more answers.

To do all that, you can’t rely solely on ETL. It’s too cumbersome and inflexible. Instead, you need a unified platform such as Boomi that lets you build integrations in just days or hours. With Boomi, you get the business insights you need – fast.

In fact, some young and cloud-centric companies are skipping ETL altogether and going straight to iPaaS. These companies find iPaaS suits their need for moving data and getting answers quickly.

To be sure, ETL will remain an important technology for large companies that process transaction data and other types of structured data as part of repeatable processes. But iPaaS is the right technology for many other types of data operations, especially those at cloud-centric businesses. For them, moving and analyzing data requires speed, flexibility and agility. And iPaaS is an essential part of their IT architecture.
When migrating to the cloud, don’t just “lift and shift.” Determine what business results you’re really trying to achieve. Then work with an enterprise architect to design the cloud services that will deliver those results.

Oliver Asmus, Data & Analytics Lead, Slalom
Chapter Three

Data Wrangling: Ensure Consistency Across Applications and Cloud Platforms

"Data wrangling," or transforming data for the applications using it, is an essential skill for IT teams migrating applications — and Boomi and Slalom can help.

By Shane Fisher, Solutions Principal, Business Applications and Integration, Slalom

At Slalom we work with lots of organizations that are migrating from on-premises operations to the cloud. It’s tempting to think of these migration projects strictly in terms of applications: You start with a legacy on-premises application, then you replace it with a fast, new and scalable cloud application.

To be sure, that’s part of the work. But another important part involves paying attention to the data being used by those applications.

Later in this eBook, I’ll write about the importance of multimastering data, which is about synchronizing data across multiple applications by using a centralized management tool such as Boomi Master Data Hub. But before you can master your data, you first need to wrangle it.
Data Wrangling and Why It Matters

Data wrangling means transforming data from one format to another, typically to make the data more suitable for downstream uses such as analytics. Developers who build these transformation processes are known as “data wranglers.”

Why is data wrangling so important?

In our data-migration work, we often find that two applications need to share data but use different formats for the same data fields. For example, a financial application and a sales CRM might use different data types for locations. Now imagine that the sales department needs to record a sale. To correctly trigger an invoice in the financial app, the two applications will need to work together, ensuring that they both handle the location data in a consistent way.

In situations like this, there are several ways to translate or “wrangle” the data between applications:

- Configure duplicate data fields
- Create a table in a SQL database
- Employ Boomi Crossref tables
- Use a central data hub

Let’s examine each of these.
APPROACH ONE

Configure duplicate data fields

If only two applications are involved, you can configure each to store two data fields for each location. That way, one field can easily be translated to the other. This approach has the advantage of being straightforward. You’ll work only with the applications storing the data, and you won’t have to deploy any new applications or services.

The downside? Poor scalability. This approach tends to work well only when just two applications are involved. If five or 10 applications use different data types for the same data field, storing, updating, and retrieving all these data translations in application logic can become slow and unwieldy.
For a more scalable approach to data wrangling, set up a table storing all the relationships in a centralized SQL database. A SQL database can perform translations among multiple applications far more easily than can applications storing multiple values for data fields. When an application needs to translate fields, it queries the database, gets a response and then uses the translated value in the response to continue with its work.

To do high-volume cross-reference checks as part of an integration or automation process, you can use a Boomi database connector. It will connect the applications to the central SQL database. There, cross-reference checks can be performed at high volume.

We find that data wrangling with a central SQL database works best for organizations with strong in-house SQL programming skills. If that sounds like your organization, this could be a good approach.

The downside? For some organizations, database performance can become an issue. If you don’t have a lot of in-house SQL expertise, you’re probably going to have a hard time setting up the cache and tuning the database. And you won’t get the needed performance at scale.
Another approach to data wrangling is to leverage Boomi Crossref tables.

Crossref is a capability of the Boomi unified platform that lets you store data relationships and perform lookups quickly. A big plus is that you can set up a Crossref without any programming. And it will run faster than a database lookup.

In your integration mappings, you can cache Crossref table lookups. That way, you'll get the benefit of having a centralized lookup, but without the huge performance hit that sometimes results from working with SQL.

For example, using our location data scenario, you could look up a country code that comes as part of an integration, get the key for the country code, get its ID, and cache that relationship. Then, the next time the same ID comes through an integration being managed in Boomi, you'll already have its relationship available for use. Translations run quickly, and multiple applications can easily be provided with the data translations they need.
APPRAoch Four

Use a central data hub

To handle data transformations and complex business logic for your most important data types, consider using a governance tool such as the Boomi Master Data Hub.

Boomi Master Data Hub can store data relationships, handle complex business logic, and enforce master data policies as needed for data types and data fields. Because the Master Data Hub is centralized and part of the unified Boomi Platform, it’s easy to access while mapping integrations and data transformations.

Boomi Master Data Hub is especially useful for organizations that need to automatically detect changes in data sets and enforce “survivorship” rules. (In data management, survivorship means systematically dealing with cases where data is inconsistent or overlapping.) The Boomi Hub is also a great choice for data wrangling when organizations want to guarantee data quality for key data types, typically as part of data-integration and application-modernization projects.

Boomi is also a great choice when having clean, accurate data is essential. Data sources aren’t always reliable; think of any task involving manual data entry, or data that resides in two different systems. With the Boomi Master Data Hub, you can first define data-quality rules and data-enrichment services (for example, external address validation services), and then catch and correct bad data before it spreads to other systems.
Choose the data-wrangling approach that works best for you

As you can see from this list, when it comes to wrangling data, you have several choices. Which approach is best? The one that best meets your needs for ease of use and scalability.

By wrangling your data, you’ll ensure that the applications you’re integrating can translate data from other applications to preserve the data’s accuracy and consistency. That can help you get the most from both your data and your cloud-migration investment.

To learn more about using the Boomi unified platform for data wrangling and multimastering data, contact Slalom or reach out to a Boomi integration specialist today.
Chapter Four
Give Your Business a Lift with ‘Data Gravity’

By making the most of “data gravity,” smart businesses can get the data and insights they need to reach new heights.

By Oliver Asmus, data & analytics leader, Slalom

IT managers and data scientists have something in common with pole vaulters: gravity. Pole vaulters need an intuitive sense of this strong, invisible force; it’s what they’re fighting against. Similarly, IT managers and data scientists need to develop an intuitive sense of a powerful force known as “data gravity.”

The term “data gravity” was coined by Dave McCrory, then a cloud architect for Dell, in a 2010 blog post. His basic idea: The more data that collects in a single place, the more valuable it becomes; this also encourages users to move other data sets to that same location, as well as to move applications and services closer.

You might think enterprise IT has become more distributed than ever. After all, more devices, services, and storage are moving up to the cloud and out to the network edge. But at the same time, data gravity is pulling things in the other direction.

Data gravity, McCrory wrote, has “the same effect gravity has on objects around a planet.” While he used the concept to explain the product launch of a new data service, it can explain a lot of other things, too, including IT development strategies and overlooked opportunities for data analytics. Find a center of data gravity, and you’ve found a place with the answers to many of your toughest questions.
Data gravity in action

To understand how data gravity can help your IT organization, imagine a financial-services company that uses Hadoop to collect data into a data lake. Because this data lake accepts a wide range of data formats and can store a combination of structured, unstructured and semi-structured data sets, the company finds it easy to upload data there. So as time passes, the company puts more and more data in the data lake.

Originally, all the data may have come from a single source — say, data about customer behavior from a marketing automation system. But as the data lake grows, accumulating many terabytes of data, it also becomes more valuable. As a result, people start adding data from other sources. That might include data from a marketing automation platform such as Eloqua. Then they might add sales data from Salesforce and finance data from Workday. As the force of data gravity pulls this information together, the total data set becomes richer and more diverse.

At this juncture, when an employee of the company wants to analyze data collected by marketing or from Workday or Salesforce, they know where to find it: in the data lake.

Soon, people across the organization start building applications, services and reports to work with the data lake, especially if they want to correlate data from different sources. In essence, the Hadoop repository’s data gravity pulls an ever-increasing volume of data, as well as analytics about that fast-growing volume of data.

A large organization can easily end up with multiple centers of gravity, due to the large data volume, various applications and databases, and the differing needs of various internal departments. As each of these centers of gravity becomes more valuable over time, they will also attract more applications and services around them.

A center of gravity needn’t be on-premises. It could also be in the cloud. For example, several of our clients have set up Snowflake cloud data warehouses to aggregate data quickly for analysis. In general, an organization can take advantage of whichever location works best.
How data gravity benefits from low-code integration

This is where low-code integration comes in. Using a cloud-native platform such as Boomi, you can connect data quickly. You won’t have to spend months hand-coding integrations; instead, you can use a low-code interface to build the connections in days, even hours. Integration is no longer a barrier to getting the data you want.

Low-code integration can help an organization make the most of its data gravity in three main ways:

#1: LOW-CODE INTEGRATION MAKES AMASSING DATA EASIER.

Low-code integration strengthens data gravity’s pull by making it easier to move data into centers of gravity quickly and easily.

Let’s go back to our Hadoop example. Imagine that the company’s marketing department wants to pull customer data from four systems: in-branch reporting, core banking, mobile location services and web analytics. With the Boomi Platform, they can easily connect all those sources to the data lake.

The Boomi Platform also offers ready-to-use connectors for popular applications and standard protocols such as FTP and JDBC, so companies can quickly set up connections and transfer whatever data they want. That makes it easy to grow the amount of data stored in a data lake or any other center of gravity.

#2: LOW-CODE INTEGRATION, WHEN COMBINED WITH A DATA HUB, IMPROVES THE COLLECTED DATA’S QUALITY.

Because a platform such as Boomi runs data transformations and enforces data-quality rules, it can help an organization ensure that all data going into a center of gravity is both accurate and consistent.

Developers can not only quickly build data transformations in Boomi’s low-code environment, adjusting data formats as needed, but they can also use Boomi Master Data Hub to govern data quality for all applications connecting to a data store. In this way, they can ensure that data fields and data types are consistent and accurate across all applications connected to the hub.

#3: LOW-CODE INTEGRATION STREAMLINES ACCESS TO REPORTING, AI FRAMEWORKS, AND OTHER DATA SERVICES.

Businesses can also take advantage of low-code integration to connect their massive data stores with reporting tools, AI frameworks and business-intelligence applications. Then they can analyze or report on this information.

Just as Boomi simplifies the movement of data into the center of gravity, it also helps with pulling data out and connecting it to the applications and services that businesses have set up to take advantage of data gravity. The result: superior analytics based on comprehensive data.
CONCLUSION

Get answers faster with low-code integration and data gravity

At most businesses, employees know who the company experts are. These people possess valuable information and have the “inside scoop” – even if they aren’t high on the org chart.

Data gravity works similarly. The center of data gravity becomes a source of answers for all kinds of questions, including customer behavior, sales trends and financial results. Just like that hidden expert in your company, data gravity works even when it resides in a lowly database or data lake, rather than in an officially sanctioned business application.

These centers of data gravity can be valuable to enterprise architects, IT engineers, business analysts and others. All can find these centers of gravity and make the most of them with data connections managed by the Boomi Platform.

As a true iPaaS, Boomi can help to make your centers of data gravity even bigger and more powerful. Business analysts, data scientists and department leaders can derive the most benefit from concentrations of information, wherever they are, whether on-premises or in the cloud.

Centers of data gravity accrue naturally through the habits and inclinations of an organization’s users, and they’re bound to appear in any large organization. With the Boomi Platform, you can control these forces of attraction and quickly reach new heights.
Chapter Five
7 Best Practices for Building Data Quality Into Integration Processes

Adopt these practices to enhance your cloud-migration projects.

By Michael Lien, client service lead, Slalom

Every organization runs on data. But for an organization to run well, its data must be complete, accurate and delivered on time.

Integration connects data sources to data targets so that data can be delivered on time to people, applications and devices. So much for data's timeliness; what about its accuracy and completeness? To achieve those goals, you'll need both data management and data governance.

Starting about a decade ago, many companies embarked on ambitious master data management projects, hoping to identify sources and formats for their data fields. They also sought to make their data consistent across the organization. But more often than not, these projects failed to deliver the comprehensive data quality they promised. As a result, many IT organizations became wary of any project promising comprehensive data quality.

Fortunately, IT organizations can achieve accuracy and completeness, and they can do so without undertaking vast, all-or-nothing data-management initiatives. Instead, they can build data management into their integration projects as they go.

Working with our clients, we've seen that data quality can be tackled incrementally in new integrations and data-migration projects using Boomi's integration platform as a service (iPaaS). The results include improved data quality and data governance. These projects tend to be highly focused and kept within budget, and their user communities appreciate tangible, readily apparent improvements to data accuracy and completeness.

To help your organization achieve these same improvements, here are seven best practices for building data management and data governance into your data-integration projects.
Make data management part of a bigger project

Don’t try to tackle data management on its own. People still remember those failed grandiose data-management projects from a decade ago. And because they tend to suspect projects that promise too much, they may argue against funding a comprehensive data-management project, recommending other projects instead.

A better approach is to make data management part of another IT project. For example, let’s say an organization is going to replace its on-premises CRM system with Salesforce. In this case, IT could set the goal of improving the quality of CRM data as part of the data-migration process. That way, the project would involve not only adopting Salesforce as the company’s CRM, but also ensuring that the data residing in Salesforce is more accurate and complete than the data in the current CRM. Because that kind of project scope makes sense to most people, it’s likely to gain support.

So whenever possible, make data management part of a larger project, ideally one that involves integrating or transforming data.

The Boomi Platform makes it easy to combine these two types of work. You can use Boomi’s iPaaS to first integrate, migrate and transform data. Then you can use Boomi Master Data Hub to implement data-governance rules for the data being migrated and transformed. Because all this work is handled by a single platform, implementation and management are easy.
NUMBER TWO

Start small: Tackle one type of data – or data used for a single operation or application

This best practice follows from the previous one. Don’t try to implement data quality for all your data at once. Instead, focus on either a single type of data — such as customer data — or data for a single application, such as Salesforce, NetSuite or Workday. Tackle that small project, and the organization will see that data management is achievable. Then they’ll trust your team when you tackle the next project, and the one after that.

It’s important to demonstrate success, especially early on and especially if data management has been a challenge for your organization in the past. Demonstrating that you can put a process in place for cleaning data, resolving discrepancies, and making life easier for end users will help you win the support of multiple departments. You’ll need that support before you tackle data management, which often involves more difficult decisions and higher costs.

Changing a data structure or source system is never free. So before you take on big changes that will affect lots of people and mission-critical operations, make sure you have that important buy-in.
It’s important to establish measurable goals for your data-management project. We all want data to be complete and accurate. But how do you measure completeness and accuracy?

To begin, ask yourself: For the type of data you’ve decided to manage, what would constitute data quality?

If it’s customer data, then you might want to make sure it’s complete. For example, that might mean that all customer addresses are current. And if a customer has multiple sites, that they’re all listed somewhere in your CRM system, with the primary site clearly identified. You might also need to ensure that all communications with sites associated with the customer are logged into the CRM system and kept up-to-date.

Usually, such goals cover not just the data itself, but also processes to ensure that your data is complete and accurate. So set metrics that can monitor your data-quality goals. For example, you’ll need to know whether people are logging into whatever tool you’re using for data management. Similarly, are they responding to queries? Are they resolving discrepancies? What percentage of discrepancies are being cleared in a given week? And how are your data-management processes performing over time?

Once you’ve established accountability for data quality, and once that accountability becomes reported and widely visible, you’ll find that people participate more willingly. Most of us would rather fix a problem the first time than be asked why we haven’t fixed it the third!

Monitor data management in whatever tool you’re using to measure other business results. For example, if you’re monitoring Key Performance Indicators (KPIs) in Tableau or QlikView, then use that tool for monitoring data quality as well. Data quality is one key metric you should always keep in mind.
Build data management into the integrations connecting your applications and services

Most organizations no longer centralize their data in a single ERP application the way they might have 10 or 20 years ago. Today, that data is instead distributed across a mix of on-premises systems and multiple cloud platforms and applications. This level of diversification is here to stay. In fact, you can expect most organizations to shift more of their work to the cloud over time. You should also expect cloud services and applications to continue to evolve.

What does this mean for data management? First, that your data needs to be federated into multiple systems. Because the same data has multiple uses, you might have different data stewards for different elements within the same data record. Tackling all this at once as a monolithic system is simply impractical. Instead, have data management tightly coupled into integration processes designed to address data-governance requirements for specific types of data in specific applications.

You need a solution that you can easily tailor to specific use cases. You also need a solution your business users will actually adopt without too much overhead.
See data management as an ongoing process; know that the data will change

Data is always in flux, so think of data management and data governance as ongoing work. The days are long gone when these projects could be short-term “once and done.”

Because data management is ongoing, it’s best to:

**Set your team’s expectations** about their involvement in data-management practices over the long term. To varying degrees, people’s jobs will change. That could entail just responding to an email to clear up some data discrepancies. Or it might mean attending a quarterly meeting to review different sorts of data types. Regardless, make sure people know that the project will continue to involve them, even if only for a few moments each day.

**Set the organization’s expectations** about budget and timing. Some amount of time and money will need to be invested in data management, and investments will need to be made on an ongoing basis.

**Select a data-management platform** that makes it easy to include stakeholders in ongoing work. Ideally, it won’t require stakeholders to learn new technical skills or adopt cumbersome processes.
Involve the business in your data-management project from the start

When we find inconsistencies between data records or how different departments are using the same data field, the issue is most often a business issue, not one of technology. And to resolve a business problem, you need business people. Don’t limit the project to the IT organization.

As a best practice, be sure to include business stakeholders in data management from the very start, including the project’s initial planning. Tap into the business staff’s expertise. They know what data they need, and they know how they use it. The business should also be able to tell you what kinds of benefits they expect from improved data quality, how they hope higher data quality will improve their processes and results. Be sure to build their goals into your project.

When you design processes for monitoring and managing data quality over time, involve business people there, too. Figure out who can act as an authority or data field for each record or field. Then let them contribute their expertise in ways they find convenient.
Build workflow automation into your data-management process

To give people the ability to resolve data discrepancies and data omissions quickly and easily, you’ll need to slip data-management tasks into their daily work. Don’t require people to log into a special application or portal, or you’ll probably have only limited success. Instead, use workflow automation to bring data-management practices into people’s everyday work. Then you’ll do fine.

For several of our clients, we’re using workflow-automation software – Boomi Flow – that’s tightly coupled with our integration platform. When an integration process flags a data discrepancy, the record is put into quarantine, and an email is automatically sent to the data steward or subject matter expert for that record. The email is short and to the point. A business user can see at once what the discrepancy is; then they can resolve it with just a few clicks. It’s simple in part because we hide all the process’s business logic and complexity.

With this type of workflow automation, you can bring data management to your business users, rather than requiring them to adopt new IT tools or procedures. The new process is easy and straightforward, making it likely to be adopted. That’s essential to data-management success.
Achievable data quality

By following these seven best practices, your organization can improve its data quality without embarking on a huge and expensive data-management project. The key to success, as these best practices make clear, has as much to do with people skills and project management as it does with technology. In summary:

Take an agile approach, building data management into the data-integration and data-migration projects you plan to tackle next.

Involve business stakeholders from the start, and let them have a voice in identifying the project’s outcome.

Use a technology platform such as Boomi that combines capabilities for data integration, data management and workflow automation. This will help you to build data quality and data management into your connections for applications, people, data and devices.

Remember that people will need to work differently (not harder) after you’re done, as data management becomes part of everyday jobs.
Chapter Six
Multimastering Data for Application Migration

The benefits of a central data hub for streamlining application migration includes lower development and testing costs, and higher data quality.

By Shane Fisher, solution principal, business applications and integration, Slalom

At Slalom we work with many companies migrating from legacy on-premises applications to new applications in the cloud. Many of these companies have over time built custom applications on top of their legacy applications. When the time comes to migrate to the cloud, they naturally need these custom applications to keep on working.

One way to ensure this is with multimastering data. An approach that synchronizes data across multiple applications, multimastering data turns out to be a critically important capability for complex integration scenarios.
Addressing the challenges of master data management

The scope of a cloud-migration project changes when it also includes support of a company’s custom applications. Now the organization is no longer moving data from just one business application to another. It is also building new integrations to all those custom applications. That way, regardless of whether these integrations are exchanging data with older legacy applications or newer cloud applications, they’ll continue to work.

This type of project can be approached in a few different ways. A decade ago, a company’s IT organization might have solved the problem with a massive investment in Master Data Management (MDM), synchronizing and correcting all data across the organization in one fell swoop. But many MDM projects became more expensive than anticipated, and they rarely turned out well.

Another approach is to hand-code a bunch of new point-to-point integrations that connect the older custom applications to the new business applications. But building and testing all those custom connections will take time. What’s more, those hand-coded connections will likely require ongoing maintenance, burdening the IT organization with a new and continual chore.

A better approach is to first connect the custom applications and services to a central data hub. Then you can connect that hub to both new and old applications. The hub ensures that all applications and services get the data they need; it also ensures that the data is consistent across all applications. At the same time, the hub spares companies the trouble of hand-coding lots of point-to-point connections. The organization gains broad connectivity – without the effort of point-to-point and without the expense of MDM.
Multimastering data for data synchronization

This approach of connecting one application to multiple other applications or services is known as multimastering or multi-synchronization. Multimastering involves synchronizing master data across multiple applications.

A migration project will go much better if you multimaster your data. Plus, you’ll gain the ongoing benefit of even faster and more efficient migration projects in the future.

To be clear, setting up the data hub will take some time, probably about as much time as you’d need to build the point-to-point connections for a single data object. But once you’re synchronizing more than one data object per application, the data hub investment should really pay off. You’ll get consistent data, but without excessive costs or maintenance work.

We find that for each data object being synchronized, multimastering data through a central data hub like Boomi Master Data Hub can save a couple of weeks of work. That adds up quickly. If you’re working with a big business-critical application such as an enterprise resource planning (ERP) system, and you’re integrating five or 10 major data objects, then the data hub could save you months of work.

What’s more, once you have the hub in place for multimastering data from one application, you can then use it for multimastering data from others. The data hub becomes your central point of connection and control. It’s where you integrate and synchronize all business-critical applications.

Reducing development and testing time isn’t the only benefit, either. The hub also makes integrations easier to manage and monitor. Data quality improves, too. Data fields and data records are consistent across applications and departments because the data first passes through the hub, which enforces data-quality rules on all connections. Overall, a data hub turns out to be one practical, affordable and effective approach to data management.
CONCLUSION

The benefits of a data hub for multimastering data

To sum up, using a central data hub to multimaster data can help your company by:

• Ensuring that custom applications and services keep working when data is being migrated from one application to another.
• Reducing development and testing time.
• Simplifying the ongoing maintenance and monitoring of integrations.
• Improving data quality and ensuring data consistency across both applications and services.

Whether you’re connecting off-the-shelf applications, custom applications, or both, a central data hub such as Boomi Master Data Hub can simplify your integration work while providing a convenient means of enforcing rules for data quality and data consistency.
The Big Picture
Data Management Done Right

In this eBook, we’ve surveyed several key issues of data migration and data management.

We’ve talked about the importance of focusing on data first when migrating to the cloud. Businesses should think about how they can make the most of their data. They should also consider how migrating to the cloud creates new possibilities for leveraging data in powerful ways.

We compared ETL with faster, more efficient models made possible by iPaaS. While ETL still helps many IT organizations, for fast-moving industries and use cases, iPaaS is the better solution. With iPaaS, data moves quickly, easily, and flexibly, accelerating business operations and business results.

We discussed data wrangling, too. It’s the art of manipulating data to ensure data quality and give every application and service the data it needs in the form it can use. Products such as Boomi Master Data Hub make data wrangling fast and easy.

As businesses move their data to new applications and locations, they need to consider data gravity. Data repositories with strong gravitational pull can offer new opportunities for data analysis and reporting.

We also discussed best practices for data management and governance. Businesses migrating to the cloud must pay attention to data quality, and that includes completeness, accuracy and consistency. A cloud migration offers new opportunities for reviewing data-management practices, making them more effective and, of course, cloud-ready.

Finally, we discussed multimastering data to ensure that data is consistent across different applications and data repositories, whether in the cloud or on-premises. The key here is to find the best scalable approach that works with the data architecture being adopted.

Slalom and Boomi work together to deliver transformative data migration and data management solutions to businesses of all kinds.

Take advantage of our expertise and the Boomi unified platform: https://boomi.com/ or https://www.slalom.com/
About Slalom

Slalom is a modern consulting firm focused on strategy, technology and business transformation. Operating in more than 30 markets across the United States, the U.K. and Canada, Slalom’s teams have the autonomy to move fast and do what’s right. They’re backed by regional innovation hubs, a global culture of collaboration, and partnerships with the world’s top technology providers. Founded in 2001 and headquartered in Seattle, Slalom has grown organically to more than 7,500 employees. In 2019 Slalom was named one of Fortune’s 100 Best Companies to Work For, and the firm is regularly recognized by its own employees as a best place to work. Learn more at slalom.com.

About Boomi

Boomi, a Dell Technologies business, quickly and easily unites everything in your digital ecosystem so you can achieve better business outcomes, faster. Boomi’s intelligent, flexible and scalable platform accelerates your business results by linking your data, systems, applications, processes and people. Harnessing the power of the cloud to unify everything inside and outside of a business, Boomi gives more than 9,000 organizations the ability to future-proof their applications strategy. For more information, visit https://www.boomi.com.