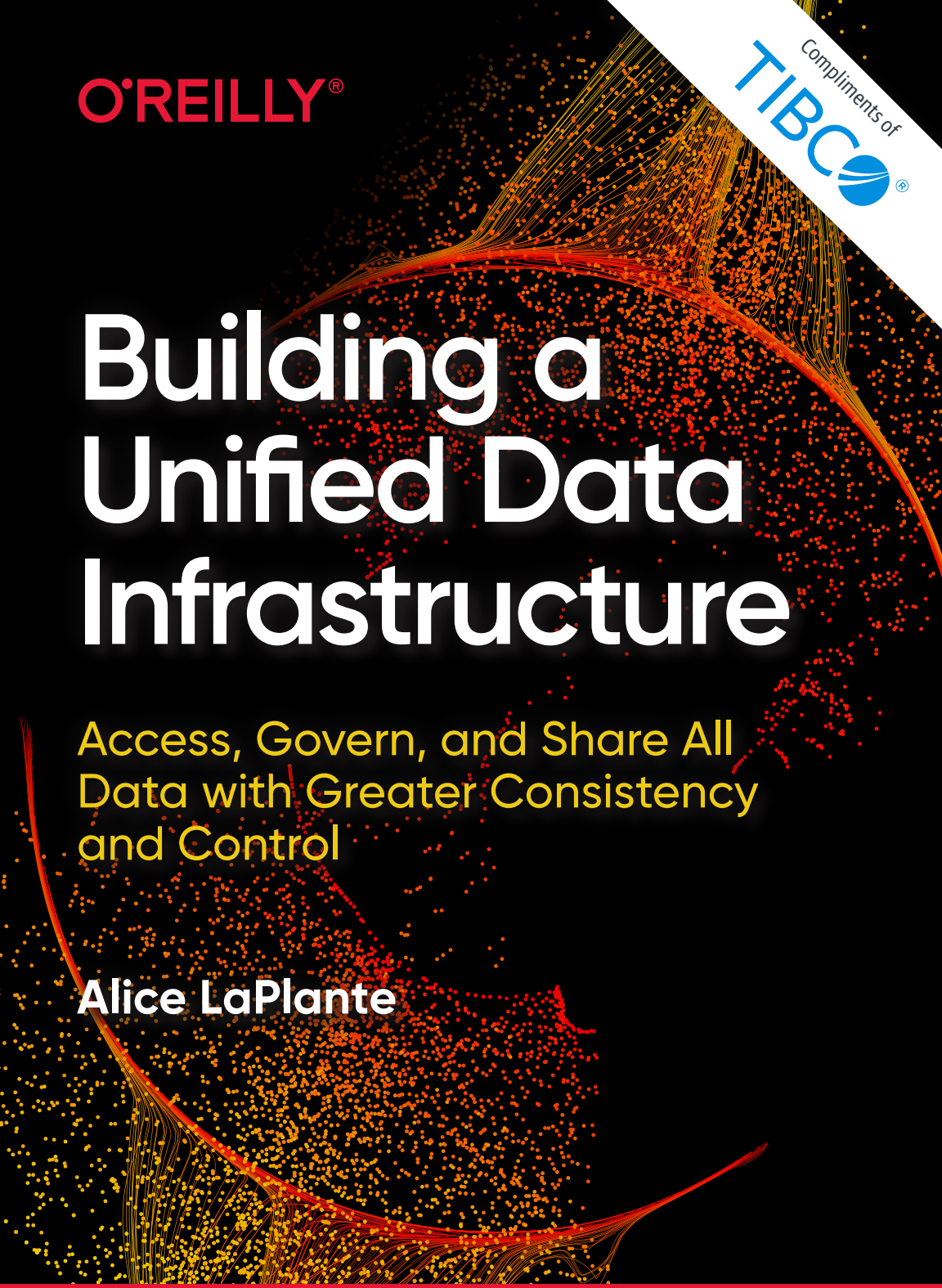

O'REILLY®

Compliments of TIBCO®

Building a Unified Data Infrastructure

Access, Govern, and Share All Data with Greater Consistency and Control

Alice LaPlante

REPORT

ADDRESS EVERY KIND OF DATA IN YOUR ECOSYSTEM IN A CONNECTED WAY

LEARN MORE

TIBCO®

Building a Unified Data Infrastructure

Access, Govern, and Share All Data with Greater Consistency and Control

Alice LaPlante

Beijing • Boston • Farnham • Sebastopol • Tokyo O'REILLY®

Building a Unified Data Infrastructure
by Alice LaPlante

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*http://oreilly.com*). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

Acquisitions Editor: Jessica Haberman
Development Editor: Corbin Collins
Production Editor: Christopher Faucher
Copyeditor: Holly Bauer Forsyth
Proofreader: Rachel Head
Interior Designer: David Futato
Cover Designer: Karen Montgomery
Illustrator: Rebecca Demarest

March 2020: First Edition

Revision History for the First Edition
2020-03-06: First Release

This work is part of a collaboration between O'Reilly and TIBCO Software, Inc. See our statement of editorial independence.

978-1-492-06322-3

[LSI]

Table of Contents

Building a Unified Data Infrastructure

Today, most businesses aspire to be insight-driven. A case in point: virtually all enterprises (97%) these days have a documented data strategy. Although many different business objectives are driving these strategies, the objectives most often mentioned are improving the decision making of business users, uncovering customer preferences and patterns, and increasing operational efficiency.

Yet only 31% of businesses claim to have managed to evolve into "data-driven organizations," and only 28% say they have built "data cultures." On every metric except driving innovation with data, firms ranked themselves as failing to transform their businesses:

- 71.7% of firms report that they have yet to forge a data culture.
- 69.0% of firms report that they have not created a data-driven organization.
- 53.1% of firms state they are not yet treating data as a business asset.
- 52.4% of firms declare they are not competing on data and analytics.

Companies are gathering and storing *a lot* of data, but the biggest obstacle to becoming data-driven is that they lack the right software tools to collect and synthesize this data, according to Prevedere's 2019 Executive Survey.

Capturing value from data requires excellence in your operating model, or the people, processes, and systems used to manage your

data. Within systems, this means having access to the right *data infrastructure*.

Data infrastructure is a digital infrastructure that enables seamless sharing and consumption of data. Similar to other types of infrastructures, it provides the structure needed for an organization to operate in a data-centric economy.

This report aims to help chief data officers, enterprise architects, and line-of-business (LOB) executives learn the importance of taking a unified and holistic approach to data infrastructure. It highlights the ways in which business objectives are met by the following solutions:

- Data virtualization (DV)
- Master data management (MDM)
- Reference data management (RDM)

The report discusses these solutions, explains why you need them, and explores the benefits of combining them. It also shares best practices on how to build a unified data infrastructure using all these technologies.

Data Must Support Operations, Analytics, and Governance

The first challenge you face when building a robust and effective data infrastructure is that there are three different contexts in which data is used within an organization:

- Operations (running the business)
- Analytics (understanding the business)
- Governance (controlling the business)

Each of these areas has its own unique data requirements and challenges, yet you somehow have to build a data infrastructure that fulfills the needs of all of them.

Operations

The operations domain includes anything that involves the running of your business. This encompasses everything from ordering raw

materials, to making goods, to processing customer orders, to distributing products, to after-sales support.

The key challenge with modern enterprises is that many of them have complex and siloed information systems. Data is duplicated and used throughout operations. While these silos are often appropriate—after all, these systems solve different problems—there needs to be a single, central point of truth for the data that should be consistent *and shared* across all the systems.

The biggest data challenge for operations is data fragmentation. Chances are good that your organization has set up multiple databases or even data lakes and almost certainly has deployed multiple applications, each of which supports a different aspect of operations and each of which generates data in its own right. So right away, you face the challenge of *accessing* that data and finding a way to get a transparent and holistic view of it—a view that represents that proverbial single source of truth.

For example, customers of banks commonly request overviews of current activities and balances in their various accounts. Although this may sound simple enough, underneath the covers, savings accounts may be in one system, checking accounts in another, and other banking products and services in others. Yet customers want to see the whole picture when they log into online banking. They don't want to have to enter usernames and passwords for multiple systems.

To solve this and similar challenges, many organizations are increasingly moving data to the cloud, where they can then expose it to whomever needs it. But some data can't get moved to the cloud, either because it becomes too expensive—cloud companies charge you by volume—or because of regulatory concerns. This leads to "hybrid" environments where some data is in the cloud and some in on-premise systems. Straddling that dichotomy presents even more access challenges.

Consistency is another critical issue when it comes to operational data. The enterprise dataset has to support all operations, from supply chain management to customer relationship management to post-sales support. If the data isn't described appropriately using metadata—data about the data—and if it isn't measured using the same metrics, you run into problems.

For example, in a manufacturing firm, starting with production, you need access to accurate raw materials data. You need to make sure that that data is appropriately and consistently named and uses the same units of measurement as it moves from the raw materials to the manufacturing database—and so on, all the way through sales and support.

The problem with not having a single view of product data means you might manufacture a certain widget identified one way, yet sell it under a different identification number. If you attempt to reconcile sales with production, you end up with "air holes" in your data.

Another increasingly common challenge with operational data is that it is a mix of transactional data and data from the accelerating growth of the Internet of Things (IoT). So, you may have sales data from your point-of-sale database, but also data coming from production oil wells, or climate controls of campus buildings, or servers on your network.

Capturing and using this streaming data in motion from your maintenance management systems, logistics systems, and planning and control systems so you can optimize your operations is increasingly an issue. You need a data infrastructure that can deal with both the highly distributed data coming from transactional systems and the real-time IoT data coming from remote or mobile devices, and that can also provide the definitional details, or metadata, such as units of measure.

Analytics

The analytics data context includes anything that involves using data to inform (i.e., guide) business decisions. This encompasses everything from factory supervisors deciding how many widgets to manufacture next month, to marketing executives figuring out how to price them, to CEOs determining optimal markets in which to expand distribution.

These particular usages of data raise specific data infrastructure requirements and challenges. First, the data silo problem is the same as it is with operations. Whether for business intelligence and reporting, self-service analytics, or data science experiments, getting access to all the data needed to perform a particular analysis can frequently be a struggle. The problem, again, is that the necessary data resides in multiple systems.

For example, customer demographic information may be in one system, but the point-of-sale transactional data—that is, what individual customers are buying—can be in another. If you want to analyze which marketing campaigns customers are responding to, or customer lifetime value scores, or a particular customer's propensity to churn to a competitor, you need access to a broad array of systems.

Data syntax and naming conventions must also be business user friendly and consistent. You can't have a customer identified as Cust_ID in one system and Ref_ID in another. Your data infrastructure must harmonize that and make it easier to identify and access the data. This is an example of what is called *metadata management* or *data cataloging*.

Data analysts also spend a great deal of time on data preparation, with some estimating as much as 80% of their time is spent on deduplicating, cleansing, and ensuring the quality of data. Having each person who wants to analyze data do this data preparation themselves is not sustainable for a business. The data infrastructure must include mechanisms for transforming and cleaning the data and putting it in one logical place for analytics. This is one example of what data virtualization can enable: a single, logical location for data.

The accuracy of the overall dataset is paramount, as it is in operations. Many business activities are based upon analyzing data, from visualizations of past quarterly profit margins to predictive analytics of which customers are expected to churn or which ones will respond to a campaign. And increasingly, artificial intelligence and machine learning are using analytics to calculate things that humans were previously unable to even imagine.

Many organizations have tried to address this issue by creating large data warehouses to use as analytics data stores. The problem with that approach is that many set up multiple data warehouses, and once a data warehouse is set up, it's a laborious process to make changes or add data to it. For known data and known use cases, they work well. But with the rise of self-service analytics and ad hoc requests for data, organizations can't fulfill those requests quickly enough. The result is that the process can be cumbersome, slow, and costly. This is why many companies are turning to master data management and data virtualization.

Governance

The governance data context includes anything that involves controlling your organization—making sure that it is operating within both formal regulatory and internal standards. This includes safeguarding the data itself as well as using data to ensure that various aspects of an organization—financial, HR, sustainability, and other sensitive areas—are in compliance with established standards.

First and foremost is ensuring that the data itself is secure. Safeguarding data privacy is turning out to be a major challenge as more laws mandating that personally identifiable information (PII) is protected are passed around the world. The European Union's General Data Protection Regulation (GDPR) is one example with far-reaching consequences for businesses. California is instituting its own version, called the California Consumer Protection Act. Brazil has the Lei Geral de Proteção de Dados (LGPD), and other countries around the world are regulating what companies can and can't do with their citizens' data. There are also internal rules around who has access to what data and what people should be doing with it. This means you have to be extraordinarily vigilant about data access rights.

But this requires a balancing act. You don't want to simply lock down all your data. Without fully leveraging all the data you're collecting, your business would suffer. At the same time, it can't be a free-for-all where everyone can access everything. In such cases, you're exposing the business to the risk of heavy fines or—worst-case scenario—a data breach that makes headlines.

Secondly, data governance is also used to oversee the various activities of the company—to make sure it's operating according to its own principles for, say, sustainability initiatives or hiring inclusion goals. Financial services firms have all sorts of internal controls to ensure that customers' money is being managed responsibly. Restaurants have a unique governance challenge: if something goes wrong, and customers get sick, how do they trace back where they got a specific product or ingredient? What was the source of the illness? Having accurate, consistent data is essential for these types of investigations.

In such cases, you have a blend of tactics that are also seen in operations and analytics. How do you disambiguate the information that's coming from suppliers? How do you standardize it so you can ana-

lyze it? How can you understand all the connections in your entire value chain?

Metadata is critical for compliance. What are all the data elements you have in your IT landscape? Where are you storing PII? Having metadata that clearly labels sensitive information makes it easy during an audit of compliance checkpoints to understand precisely where it is stored and who has access to it.

Taking a Pragmatic, Holistic Approach to a Unified Data Infrastructure

As previously explained, a data infrastructure is a digital infrastructure that promotes data access, consumption, and sharing. A strong data infrastructure enhances the efficiency and productivity of both the business and technology environments in which it is employed, increasing collaboration and interoperability.

The biggest problem with building a cohesive, holistic data infrastructure that serves operations, analytics, and governance domains is that most companies have numerous different legacy databases and data stores. You might even call it technical debt: technology deployed on top of early technology on top of even earlier technology, all still working and in fact growing every month.

It's usually not possible to simply decommission your existing data infrastructure and put in a new data architecture. You have to build a bridge. In a way, it's like attempting to fix a car while driving it—and that's the challenge.

The most pragmatic journey to a unified and holistic data infrastructure takes a coexistence rather than a replacement approach. After all, you have to live in the real world, and you can't discard any of the things you currently have running for operations, analytics, or governance. But you need more flexible and efficient technology to address some of the new challenges that you're going to face with distributed computing, the cloud, the edge, and the IoT.

The good news is that the three business domains discussed in the preceding section don't require different data architectures. What you're doing for operations can be applied to analytics, which can be applied to governance, and so on. This one infrastructure allows you to embrace the chaos of data being everywhere and anywhere. By

layering on technology that provides the necessary controls, you can also get a unified picture across all the disparate data.

This cohesive data environment allows for seamless access, integration, and sharing of otherwise siloed sources.

Characteristics of a Holistic, Unified Data Infrastructure

A holistic approach to a data infrastructure includes the following characteristics:

Allows access to any kind of data
: It doesn't matter if the data is structured or unstructured, transactional, or comes in streaming form from the IoT. A holistic, unified data infrastructure can handle it all.

Enables strong data governance
: A holistic, unified data infrastructure ensures that your data is easily identifiable via metadata and that sensitive PII, Payment Card Industry (PCI), and other data can be identified and protected from unauthorized access.

Ensures data quality in six dimensions
: Even among data professionals, there is disagreement about the various dimensions of data quality. However, six dimensions are generally universally accepted:

  Completeness
  : It is inclusive of all relevant data.

  Uniqueness
  : No piece of data is recorded more than once.

  Timeliness
  : The data represents reality from a very recent point in time.

  Validity
  : The data conforms to the syntax (format, type, range) of its definition.

  Accuracy
  : The data correctly describes the "real-world" object or event under consideration.

Consistency
: There are no differences when comparing two or more representations of an object or event.

Allows data sharing across business domains with consistency and control
: This means that the data can be equally used for operations, analytics, and governance without any differences in format or accuracy.

Is able to work with existing technologies, without having to rip and replace
: This is a big requirement, as few organizations can afford to tear out all the legacy systems that generate or consume data.

Evolves as business needs change
: Technology in this area is evolving rapidly, as are businesses in increasingly digitized markets. An essential requirement is putting in a data infrastructure that can change as your business changes.

There are two technology components that enable this vision of a unified data infrastructure: master data management and data virtualization.

Master Data Management and Reference Data Management

MDM is a technology used to manage the critical data of a business and deliver a single point of reference for that data. To achieve this, MDM includes removing duplicate data, standardizing it, and imposing rules to prevent incorrect data from entering the system. The result: a master dataset that provides a single, trusted version of the truth.

MDM solutions also offer processes for collecting, aggregating, matching, consolidating, assuring the quality of, and distributing data throughout an organization. This makes sure that, throughout operations, analytics, and governance activities, everyone who touches the data shares the same understanding of the data.

A subset of MDM, *master reference data*, is also important for creating a unified data infrastructure. Master reference data is data that defines other data. It can be an external standard, such as postal or

country codes, or it can be internally defined at the company, department, or even team level. To understand the differences between reference and master data, see Table 1.

Table 1. Reference data versus master data

	Reference data	Master data
Definition	Constants that define permissible values for data	Data that has a common definition across an organization
Value	Data standardization Validation Data quality	Data quality Reduced costs Streamlining data management and governance

Panera Bread 2.0

Panera Bread Company is an American chain of bakery-café fast casual restaurants with more than 2,000 locations in North America. Its headquarters are in Sunset Hills, Missouri, and it employs more than 50,000 people.

To manage kiosk ordering, inventory, and food costs, Panera Bread needed a single source of truth about its data. In 2017, the firm launched Panera 2.0, its digital transformation initiative, which was going to digitize all the operations of the organization—managing products, menu items, café operations, and marketing. A key aspect of Panera 2.0 was building a multidomain repository of all its master data. Everything from introducing new menu items—which Panera does weekly—to opening new restaurants is updated in the master repository.

One reason Panera introduces new menu items more frequently than other restaurants is to delight its customers. Part of this strategy includes offering items that may be unique to a specific region and have an emotional connection or resonance with customers. This is why guests can find regional favorites such as the gooey butter cake in St. Louis locations, while lobster rolls appear in New England (during the summer, naturally). Better control over the data, which contributes to fewer errors in the supply chain and operational processes, makes this customer experience–enhancing variety possible.

Data Virtualization

A technology called *data virtualization* completes the picture of a unified data infrastructure. Data virtualization creates a virtual data layer that sits on top of all your data sources. It then creates a single, abstracted view of those sources. You can query this layer without knowing exactly where the information comes from or how it's formatted.

Basically, this creates a virtual data warehouse. The first benefit of data virtualization is that you don't have to move the data from any of your siloed systems into a physical data store. The second benefit is that querying can be done on demand. When you query the data virtualization layer, you are actually querying the underlying data sources. There's no intermediary data lake or data warehouse. The information, therefore, is always fresh.

By eliminating the need for a traditional extract, transform, and load (ETL) process, using data virtualization reduces the risk of data errors and the need to move data around that is never used.

Data virtualization is also noteworthy for its federation capabilities. It encompasses data across data warehouses, data marts, data lakes, and other data sources without having to create a whole new integrated physical data platform. Existing data infrastructure can continue performing its core functions while the data virtualization layer leverages the data from those sources. This aspect of data virtualization makes it complementary to all existing data sources and increases the availability and usage of enterprise data.

Data virtualization is a way to provide some agility and flexibility in responding to requests for data that come from business users. All users see is one dataset that they can access and use for their purposes. It also doesn't matter what the consuming application is. It could be a Cognos Business Intelligence report, it could be a Tableau user, or it could be a SAS or Excel user. Data virtualization abstracts away all of that underlying complexity to make the data easier to consume.

Historically, the main use case for data virtualization was analytics. These were times when you wanted to give access to data sources to business analysts and data scientists very quickly—when they didn't want to wait until you put the data in your data warehouse or data lake. Data virtualization is a great solution for those instances

because the data virtualization platform will connect to all your systems and create an abstract layer where you will only see one system.

Today, the same benefits accrue for operations. We live in a world where companies are starting to deliver their systems as a service, or as an API. Data virtualization works well for that, because you can create a layer of data APIs on top of your systems, which is an example of data-as-a-service.

Combining MDM, RDM, and Data Virtualization: Best Practices

A combination of MDM, RDM, and data virtualization creates a flexible and cost-effective data-management platform, especially when compared to the traditional approaches.

For example, say you're using an MDM solution to manage your customer data. You want a truly 360-degree view of your customers, which means you want to view all transactional data related to a particular customer. What products do they use? What's the last purchase they made? What store locations do they patronize? When you combine MDM with data virtualization, you can answer these questions in real time.

Being able to take data virtualization and infuse that with MDM and RDM components improves the consistency of and the trust in the data. This creates a cohesive environment where you can address multiple use cases at once.

When these technologies are deployed separately, any change in one of them has to be propagated to the other use, either manually or via an additional process. When they're integrated, any change in one area automatically updates the entire system, so you can move more quickly and with greater integrity.

The way it works is that MDM takes care of much of the heavy lifting of preparing the data: finding it, identifying it, cleaning it, formatting it—those activities that can otherwise eat up as much as 80% of your time, according to the *Harvard Business Review*. From an organizational perspective, that doesn't make sense, because with that type of approach—without using MDM—what you eventually find out is that, well, you're doing it, but so is every other product

manager. Everyone is cleaning up the data by hand every quarter in their own particular little silo.

Not only is this a waste of time, but there's a consistency problem. You're making different decisions from your colleagues. You might have collapsed those two deals for Company A into one deal, but another sales professional says, no, they're actually two deals, and one is a joint venture with Company B. When you have MDM and RDM in place, you can all agree on what the consistent data elements are and what the consistent entities are, and you don't have to do that work anymore.

The data virtualization part is really about the *movement* of the data. With data virtualization in place, again, somebody has gone through the trouble of pulling all the data together for you. You don't have to go into Salesforce, look for your accounts, and download the data. Instead, you just go to the data virtualization layer, and it becomes a service rather than a manual data acquisition task.

The future is one in which people don't have to do menial labor just to get to a point where they can do their jobs. They can simply perform the analyses and do the planning and strategy work they're being paid for, and have easy, transparent access to resources.

For business users, they know they have a clean, up-to-date set of customer data. They don't have to worry about getting access to transactional information because they have a data source that grabs that data for them.

IT professionals will see productivity improvements. No longer do they have to intervene when a user makes a mistake in Salesforce or changes a record that shouldn't have been changed. And these technologies add more peace of mind for IT, because there's an abstraction layer between the consumers and the producers that protects the data.

In the past, IT may have had to spend two weeks figuring out what the data source looked like, writing and testing ETL scripts, and then setting up the enterprise data warehouse. Because it's much more of a federated query, data virtualization can be set up much more quickly by fewer IT staff, which means less cost for the organization as well as better quality of life for IT professionals.

The combination of MDM, RDM, and data virtualization addresses a multitude of business problems. Being able to expose, for example,

master data through that data virtualization layer again ensures consistency and trust in the data—that it's high quality, and that the results of the analysis can be trusted. It also makes sure that that master and reference data is being utilized by the appropriate people.

Without a unified data infrastructure, there's no common identifier or consistency across your data. There are formatting differences in the data, and sometimes the data is not up to date. Say a customer moves, so their address is out of date. Or maybe they get a raise or get married. You want to make sure you're capturing these changes. By combining master data with data virtualization, you address those types of problems.

With master and reference data, you create the common identifiers and golden records needed to ensure consistency and trust and high quality. But then you can combine that data through data virtualization with on-demand federation of those different data sources, particularly from transactional data. You can see what customers are purchasing, what marketing campaigns they're responding to, what they're putting in their shopping carts. This knowledge can drive new types of analysis and be used to refine other marketing campaigns or sales efforts to make sure that you're targeting those customers who are most likely to buy or most likely to respond.

Don't Virtualize Everything

When you have known use cases and known users, then it's probably a good idea to physically integrate that data and put it into your data warehouse or data lake so users can access it.

But when you have unknown data with unknown use cases and unknown users, or when you're getting ad hoc requests for data, data virtualization makes sense. This gives you a way to manage those types of requests and provides you with flexibility, in that you've created a sandbox that users can access to get the data they need.

Implement the 3-3-3 Rule

When you get a request for data, put in a data virtualization layer within three days. Then, check back in three weeks to see if it's still being used. Is that data view or data service that you built still being accessed? If so, refine the service and keep it active. Then check back

again after three months. If it's still being used after three months, then you know it's time to put that data into your data warehouse.

One company that implemented this rule found that most of the requests—perhaps 80%—for data never made it past the three-week marker. This allowed them to get a better sense of where requests for data were coming from and what they were for. It helped them answer the question of whether they needed to go through the laborious, manual steps of putting the data into the warehouse.

Have an Enterprise Vision, but Use a Pragmatic, Incremental Approach

You need an enterprise-wide vision for your data infrastructure. But at the same time, you should possess a very pragmatic, incremental approach to delivering it. This means that, for each domain you're going to attack or each business function you're going to try and help, you need to be able to deliver value, while at the same time making progress toward your overall goal. You should be very ambitious around data management, but be patient enough to go through small wins, step by step, to get there.

This incremental approach is important for several reasons. First is cost. The problem with data management is that no one department is willing to pay for it—it benefits all, but nobody is really the owner. Second, the data landscape is so big and so vast, that if you try to do it all in one go, you're looking at a 10-year program. And from experience, you know that any project that lasts more than 18 months is going to fail.

The difficulty of change management is another reason to take small steps. People within your organization will need to change the way they work. They cannot continue to manage the data in their own silos. They will need to collaborate and adopt governance processes across the enterprise. For all those reasons, a big single program never works. Again, it's very important to set a target that is ambitious, but there should be a succession of small projects that will add up to reach the target.

Take a Model-Driven Approach to MDM

There are two different types of data-management technology. The first provides you with prebuilt data models to manage your product and customer data. The benefit of that approach is that it works out of the box. The problem is that this approach is very rigid. And in real life, nobody really fits into a prebuilt model. They always need to customize it.

The second type of data-management technology is *model-driven*. This means that what you model is what you get. A model-driven MDM solution is where you can design your own data model to define what you want to manage—for example, your own customer model and your own product model. With data virtualization, it's the same. You get a solution that can connect to any systems—databases, data lakes, or enterprise apps—to create your own data virtualization layer. This gives you maximum flexibility to adapt the data-management platform to your own requirements and your own organizational specificities.

Use the RACI Model

RACI is an acronym in a commonly used responsibility model, derived from four key obligations: responsible, accountable, consulted, and informed. It is used for clarifying and defining roles and responsibilities in cross-functional or departmental projects and processes. It's important for putting together a data infrastructure strategy because it focuses on people—the individuals who are going to be responsible and accountable for the success of the project.

That generally starts with a core team of operations people to be responsible and accountable for helping IT with the data architecture. Then, you've got to think of the wider picture. Who's going to be consuming the data or consulting with these people? And who's going to be informed only?

Sometimes you focus only on the responsible and accountable and ignore the consulted and informed. That's a mistake, because you're talking about the data that the organization as a whole is going to depend on and that it's going to share to improve transparency throughout operations, analytics, and governance. To facilitate better movement of materials through the supply chain and through

the operations chain to the customer and beyond, you need to broaden your inclusion of relevant people.

What that means is that you should have a substantial number of seasoned consulted and informed people in your RACI matrix. You should especially focus on who the consumers of the data are going to be.

Crawl, Walk, Run

Imagine a quadrant of effort along the *x*-axis and value along the *y*-axis (see Table 2).

Table 2. Quadrant of value versus effort

	Low effort	High effort
Low value		
High value	XX The sweet spot XX	

Go to that quadrant of projects that will deliver the highest return for the lowest level of effort. That way, you drive the most benefit and at the same time build up your skills and systems. Don't do the higher-effort, high-value projects until you've done the lower-effort, high-value ones.

The first step is always to get clarity about your customers and where the data about your customers resides. Those are the MDM and RDM parts. Data virtualization lets you access the details of all that data in all those different places. Then you can slowly build up, investing increasing amounts of money and effort in your data infrastructure journey.

Summary: Benefits of a Holistic, Unified Data Infrastructure

Putting a holistic, unified data infrastructure in place that integrates MDM with data virtualization drives a number of important benefits to organizations.

First, it promotes agility and flexibility to fully support all three of the main business domains discussed in this report (operations, analytics, and governance).

Second, it delivers trusted enterprise information for all data consumers throughout your enterprise. No matter what data they need to access, they can rest assured that it is up to date, consistent, and of high quality.

A holistic, unified data infrastructure also builds a cohesive data environment that allows for seamless access, integration, and sharing of otherwise siloed sources of data. And it leverages the best storage and compute resources available.

Having a unified data infrastructure provides both control of the data and visibility into where the data is, what it is, and who is doing what with what parts of it. By having it all unified, IT has the ability to provide control and protections and get better insight into data lineage.

But it's important to remember that, ultimately, this is not a technology problem. It's all about people and processes. You need to first establish a vision and a roadmap to reach that vision. What is your long-term goal? What do you want to achieve? And how can you get there in a step-by-step manner? Following the best practices set out in this report will start you on your journey.

About the Author

Alice LaPlante is an award-winning writer, editor, and teacher of writing, both fiction and nonfiction. Alice's roots are in technology journalism. She was news editor of InfoWorld for five years, and a contributing writer to other national technology publications, including ComputerWorld, CIO, and InformationWeek. A Wallace Stegner Fellow and Jones Lecturer at Stanford University, Alice taught creative writing at both Stanford and in San Francisco State's MFA program for more than 20 years. A *New York Times* bestselling author, Alice has published four novels and five nonfiction books, as well as editing bestselling books for many other writers of fiction and nonfiction. She has consulted with Silicon Valley firms such as Google, Salesforce, HP, and Cisco on their content marketing strategies. Alice lives with her family in Palo Alto, California, and Mallorca, Spain.