O'REILLY®
Business Guide

Logical Data Management

An Essential Data Strategy for Transforming Your Business

Christopher Gardner

denodo

LOGICAL DATA MANAGEMENT

Deliver Data in the Language and Speed of Business, in this Age of AI and Increased Data Complexity

The Denodo Platform enables the rapid delivery of data products through a unified business data layer with centralized security and governance.

Learn why leading companies have voted us as a Gartner® Peer Insights Customers' Choice four years in a row, with an average rating of 4.6 out of 5, as of 2024.

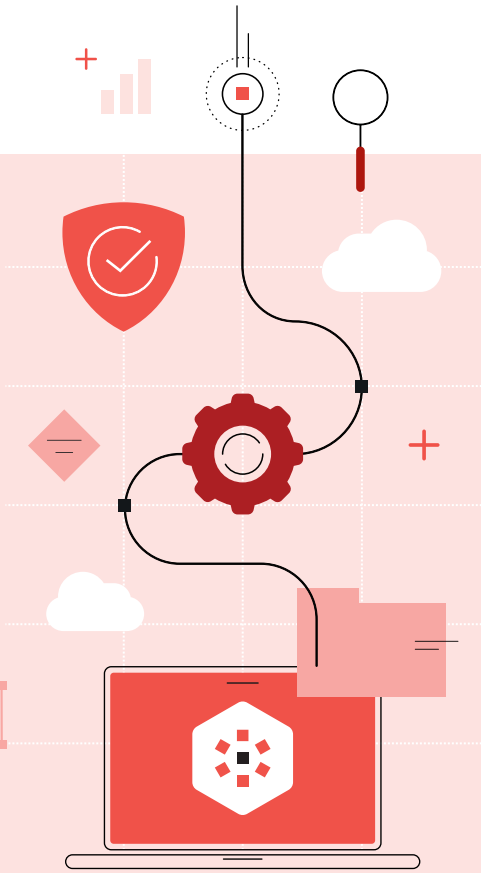
GET STARTED

Agility is only the beginning...

The Denodo Platform is smart, flexible, powerful, and affordable, to support myriad modern use cases.

Compared with traditional data management platforms, it enables:

- **65%** faster data delivery
- **67%** less data preparation
- **83%** faster time-to-value



Source: *Independent third-party analyst firm in a 2021 report on the impact of Denodo's data virtualization, a key component in logical data management.*

Logical Data Management

An Essential Data Strategy for Transforming Your Business

With Early Release ebooks, you get books in their earliest form—the author's raw and unedited content as they write—so you can take advantage of these technologies long before the official release of these titles.

Christopher Gardner

O'REILLY®

Logical Data Management
by Christopher Gardner

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*.

Editors: Virginia Wilson and Andy Kwan
Production Editor: Kristen Brown
Interior Designer: David Futato
Cover Designer: Ellie Volckhausen
Illustrator: Kate Dullea

August 2025: First Edition

Revision History for the Early Release

2024-07-15: First Release
2024-12-19: Second Release

This work is part of a collaboration between O'Reilly and Denodo. See our statement of editorial independence.

978-1-098-17604-4

[LSI]

Table of Contents

Brief Table of Contents (*Not Yet Final*)

Preface

Data management has been around as long as we have had data. To the layperson it sounds easy. You have an observation, and you record it. But as anybody who has worked with data can attest, managing data is often fraught with complexities. Figuring out how to structure data so it can be found easily and quickly, as well as keeping it accurate, is hard in itself. On top of that, the same data needs to serve the disparate needs of different users.

If you're navigating the complexities of today's data landscapes, or if you have largely simplified your landscape with a modern data lakehouse solution, you may encounter challenges delivering the right data, formatted in the right way and at the right time, to meet dynamic business needs.

Who This Is For

This book is for business leaders and senior technologists who are charting the course through today's complex data ecosystems. It provides an in-depth look at logical data management, shedding light on its inner workings and demonstrating its strategic value across various businesses. Whether your objectives involve enhancing AI capabilities, promoting data democratization, transforming customer experiences, or strengthening governance, risk, and compliance initiatives, this approach has proven successful for numerous companies. As you navigate through the upcoming pages, you'll find ample examples that illuminate the realm of possibilities and guide you towards achieving powerful results.

What You Will Learn

As you explore this book, you'll gain insights into logical data management approaches that will equip you with the knowledge to transform your organization's data strategy. Here's what you can expect to learn:

- Uncover the essence of logical data management and its transformative effect on data management.
- Learn how logical data management platforms can augment the capabilities of modern data platforms like cloud data warehouses and data lakehouses.
- Recognize traditional data management's flaws and the imperative for change.
- Learn the foundational principles and operations of logical data management.
- Observe how it collaborates with AI to jointly amplify data utility.
- Learn to broaden data access and fortify organizational decision-making.
- Utilize insights to innovate customer interactions.
- Investigate logical data management's role in redefining governance, risk, and compliance.
- Identify tactics to elevate operational efficiency and agility.
- Decode logical data management's contribution to IT infrastructure advancement.
- Look ahead to the future developments and trends in data management.

By the end of this guide you will have the tools to not just manage data, but to use it as a driving force for innovation and progress. Let's get started.

CHAPTER 1

The Challenges of Managing Data

A Note for Early Release Readers

With Early Release ebooks, you get books in their earliest form—the author's raw and unedited content as they write—so you can take advantage of these technologies long before the official release of these titles.

This will be the 1st chapter of the final book.

If you have comments about how we might improve the content and/or examples in this book, or if you notice missing material within this chapter, please reach out to the editor at *vwilson@oreilly.com*.

In the whirlwind of the 2020s, marked by unprecedented challenges and technological leaps, the quest to master data remains more critical than ever. With the rise of transformative technologies like Generative AI, the urgency for organizations to effectively manage and utilize data has intensified. Despite modern advances, distributed data continues to be a major challenge. The ability to provide actionable insights to decision-makers, crucial for navigating complex scenarios and driving success, hinges on the proficiency of data management.

Enter logical data management—a strategic approach designed to meet this need head-on as a beacon for organizations, guiding them in the art of transforming data into a language that resonates with

its users swiftly and accurately. It's not just about managing data; it's about empowering it to be a decisive tool for achieving customer satisfaction, operational excellence, and innovative breakthroughs. And it's not about replacing solutions like cloud platforms, cloud data warehouses, or data lakehouses, it's about augmenting their capabilities in powerful new ways.

But before we dive into logical data management, let's review the challenges of managing data in today's data landscape through the eyes of data analysts, who need to gather data and translate it into actionable insights for your business. This data can come from anywhere in the organization and occasionally from outside as well. Their challenge is to get access to that data, understand it, analyze it, and deliver the results to your executive team. This means they have to be knowledgeable not only in data analytics and reporting, but also in how to get access to the data, what types of questions your executive team wants answered, and what data will deliver those answers.

Transitioning From the Traditional Data Flow

How do they translate data to decisions with such a wide variety of sources? Many modern organizations rely on extracting data and loading it into a data warehouse. In this data flow, the divisions within the company produce data using transactional systems. This data is usually structured for the purpose of running the application and not for reporting. As a result, most organizations extract the data out of the transactional database and place it into a reporting database. Through a process called ETL (Extract, Transform, Load), data is drawn from the transactional systems and translated into a structure that better reflects business processes reporting needs. From there, it is loaded into a reporting database where it becomes available for analysts to develop reports.

Data Warehouses

Data warehouses contain data from all over the organization, but it is often not detailed enough for specific reporting needs. When this happens, a smaller subset database is developed called a data mart. The data mart contains a focused set of data specific to a functional section or line of business within the organization. This detailed data allows analysts to develop reports at a more granular level.

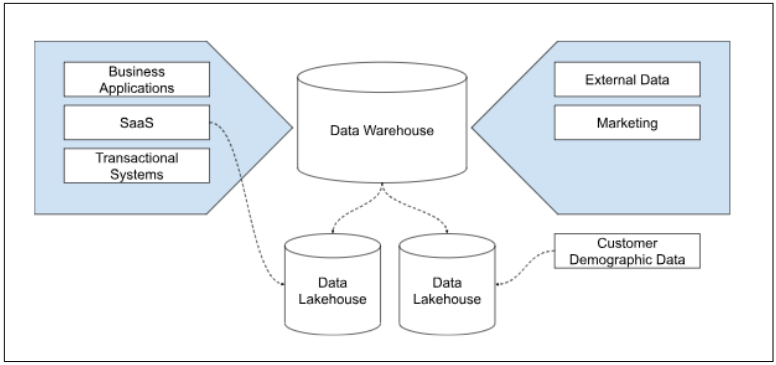


Figure 1-1. An example of how data can be collected in a data warehouse. The business collects as much information as possible and loads it into a central location, which creates a collection of schemas, tables, and potentially poor quality data. Data lakehouses (see below) make improvements on the data warehouse architecture.

Data warehouses and data marts provide a basis upon which data analysts can build reports. Until recently, those reports were one of the few ways executive leadership could gain understanding into how the company performed. The data landscape is changing, however. Data science, AI, and machine learning have transformed how organizations look at data. They've also exposed the challenges related to traditional data management.

Data warehouses have done great service to the business community, integrating data from OLTP databases and consolidating them into analytical data in an OLAP database. But one of the downsides of data warehouses is that they are physically installed on a single piece of monolithic hardware, and they are expensive to scale with more storage and computing power. If a data warehouse grows and exceeds the server capacity, the administrator has to buy a bigger and more expensive server.

On another issue, the queries that business analyst users write against an OLAP database are very different from the ones sent to an OLTP database. They are usually complex and expensive to run on the data warehouse. Data warehouse administrators have to work hard to optimize the tables and make compromises by denormalizing the data, and creating indices that speed up queries but slow down writing data to the data warehouse.

We have to consider the cumbersome and time-consuming expense of ETL jobs itself. As analysts and business users demand more "real-time" data, there is stress to increase the frequency of loading data into the data warehouse. Bottlenecks in processing time and bandwidth are encountered, and more slowdown occurs in reading and writing to the data warehouse.

Another point to consider is that with many different business users having different needs, it is easy to encounter situations where the same data has to be copied and modified into many different formats for different use cases. These *data marts* can add complexity, processing time, and expense to serve these different end users. Conversely, with ETL it can also be difficult to reuse transformation logic across multiple pipelines, so common transformation logic ends up repeated multiple times and maintenance becomes very hard.

The data landscape has been moving beyond the capabilities of the traditional data warehouse, as new technology and practices have been emerging in the marketplace with relation to data. Technologies such as Artificial Intelligence (AI) and Machine Learning are opening avenues to new data insights. The data science field is exploring how statistics can create predictive and prescriptive models for future business decisions. Data is also expanding from structured data in columns and rows to unstructured data like audio, video, and images.

How can you take advantage of these new innovations in data? How does your current data pipeline support or prohibit utilizing these resources? Most importantly, how can your company overcome existing obstacles to take advantage of the opportunities these data sets provide?

The Growth of Data

To be successful, you want data from multiple sources. This is not just the data your own organization creates during the normal process of business. This is also data from beyond the scope of business function:

Additional Internal Data
: The core functions of your business create data that is stored in the data warehouse for reporting purposes; however, there are usually smaller divisions within the organization that also

create data. For example, data might be collected from machinery during a manufacturing process to track uptime and maintenance. Without analysis, the data collected by this machine is effectively useless.

Data from Business Related Processes
: Some data comes from sources related to, but not part of, the business. These would be sources such as supplier data, marketing campaigns, and product distribution providers. These sources are not part of the main business process per-se, but the information they provide is no less valuable. They can open up opportunities for cost saving, logistical changes, and avoiding potential obstacles that may impact your business.

Data from External Sources
: Some data might seem unrelated to the business as well, but it can still have an impact. Information about road construction, weather, customer demographics, social media and more, all have impacts on the success of a business. For example, a natural disaster might disrupt suppliers within certain regions of the country. These external sources are a primary source for analysis by data scientists and AI.

Data is everywhere, especially with the advent of the Internet of Things (IoT). Objects such as watches, cars, phones, and manufacturing machinery all collect data. They provide valuable information about customer habits, hardware life, travel patterns, and more. The data they collect can provide actionable insights with regards to acquiring customers or providing adequate maintenance on a piece of equipment. There are thousands of different objects all collecting information that could be potentially vital to your business.

All of this data is not easily accessible. It comes in multiple formats and multiple structures. Companies need a way to consolidate this data to make it usable. It needs to be collected, organized, and stored for reporting purposes. It might also need to be combined with business related data to derive its impact. All of this takes time, storage space, processing power, and the employee resources to manage it all.

Data science has also influenced the evolution of data management. Data scientists look across many different domains to extract patterns, correlations, and trends. They derive predictive and prescriptive plans that help the company make decisions looking ahead

instead of reporting on what happened in the past. This analysis takes large volumes of data from many different sources.

The analysis data scientists perform takes large volumes of data from a wide variety of sources both in and outside of the organization. The variety of data creates a challenge for the business, as these data sources are likely in different formats and structures. To make them accessible and usable for the data scientists, the data needs to be translated into a common format. The data needs to be massaged and manipulated to enable it for use in analysis.

To complicate things further, data is no longer limited to a structured format. A *structured data set* is one that can be categorized and organized into a table. Its information fits neatly into rows and columns and uses discrete data types to hold its values. This is what data warehouses are built from. The data is organized in a way that can be easily pulled, categorized, and measured.

The alternative to the structured format is unstructured data. *Unstructured data* is data that doesn't follow a consistent format and usually requires much more work to extract information. Examples of unstructured data include images, videos, audio, webpages, survey data, and more. To utilize unstructured data, businesses need a place to store it and methods to extract its values such as natural language processing, machine learning, and AI. This unfortunately leads to a whole new set of challenges.

Data Lakes

For unstructured data, many organizations turn to data lakes. Data lakes are storage structures for data that maintain the information in its raw form. Data lakes can hold both structured and unstructured data, as there is no requirement to modify its format. They also alter the way data is integrated and stored by extracting and loading the data without transforming it until it's needed. This transfers the resources needed to transform the data from the extraction tool and places it on the tool that's utilizing it for analysis. Data lakes can also be the source for an ETL process that extracts the structured data from the lake, transforms it, and loads it for use in a data warehouse.

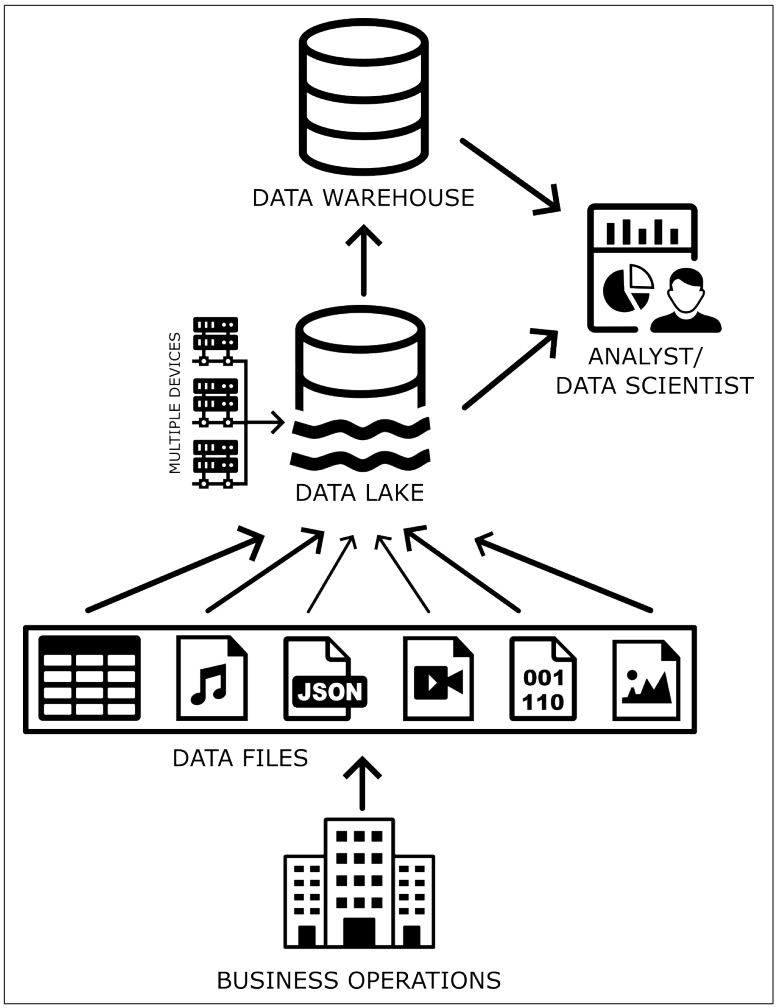


Figure 1-2. Data lake

It is probably no surprise that data lakes' greatest strength, their flexibility, is also their weakness. By making data lakes a free-for-all where any and all data can be uploaded in any format, it can become chaotic very quickly. Data lakes typically have a "schema on read" strategy, meaning that there is no enforcement for the data to follow a certain format or shape (unlike relational databases, which have strict table structures and rules). This is why if data governance is not prioritized, a data lake quickly becomes at risk of becoming a "data swamp."

Data lakes also put an emphasis on ingesting data in their raw formats straight from their sources, which helps keep data timely. This has a downside of those datasets not being processed to be user-friendly. Data lakes often are the domains of power users like data scientists and data engineers who have technical skills beyond just SQL. They also are not performant for analytical queries. This inevitably created some tension with business end users who are proficient in SQL but are not engineers or programmers. This can add processes and work that are expensive. For this reason, a data warehouse is still needed to receive cleaned and transformed data from a data lake for business end users. A strong argument can be made that this marginalizes the value of a data lake greatly, bringing all the problems of data warehouses. I will talk about how the data lakehouse attempts to address this shortcoming shortly.

Another annoyance with data lakes is they do not offer transactional guarantees like relational databases do with ACID standards (Atomicity, Consistency, Isolation, and Durability). Since data lakes make heavy use of files divided across different nodes, they can only easily append data to files. Modifying existing data is time-consuming and expensive because it would have to rewrite a whole file for a single change. The solution to this was to break up the files into even smaller files so the rewrites were minimized, but this led to the "small file" problem where there are now many small files to represent a single file. This in itself would cause performance issues to the data lake and require further administration.

Pulling back, you can look at a data lake as the solution to data silos happening across an organization, but still not addressing the problems of data warehouses and ease of use for analytics users. A data lake can work great for certain environments with only technically-proficient users who prioritize flexibility and scalability above all else. This is why data warehouses (and many alternatives like NoSQL) continue to be used. So let's talk about the next logical step, absorbing the data warehouse function into the data lake.

Data Lakehouses

In some cases, businesses may take a hybrid approach to data, storing it into a data lakehouse. A data lakehouse is a hybrid approach where data is stored in both a structured and unstructured format. The structured data is utilized by the business intelligence side of the business for analysis and reporting. The unstructured data is

utilized by data scientists through AI and machine learning. Some organizations utilize natural language or other AI tools to analyze the contents of the unstructured data in order to apply a structure or categorization to its contents.

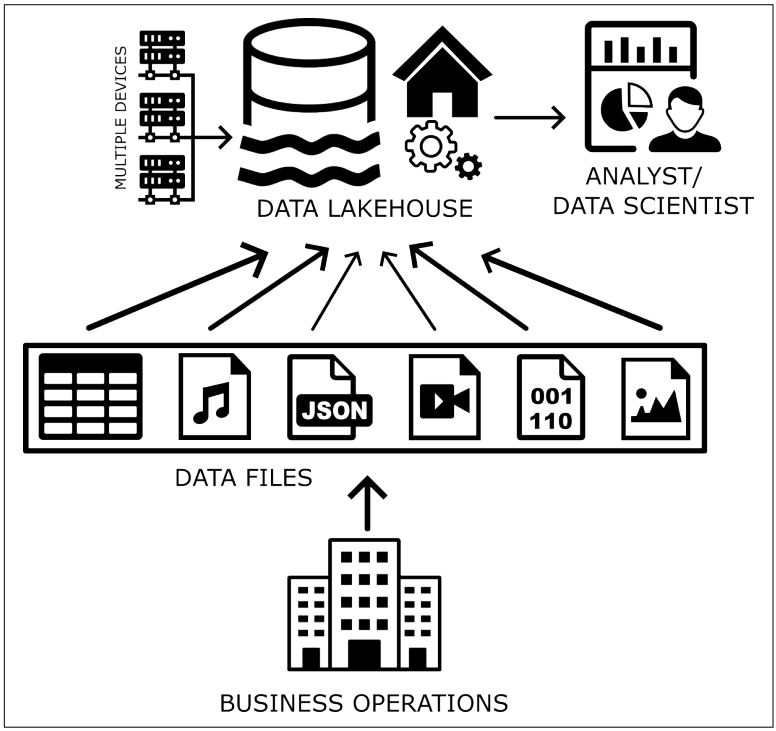


Figure 1-3. Data lakehouse

The main problem with data lakehouses is that they can never be an organization's sole repository. Some on-premises data will always have to remain on-premises, for compliance with data privacy and other regulations. Also, some organizations will want to foster and maintain a multicloud infrastructure, to leverage the best features of different cloud platforms. In addition to simply desiring to keep certain cloud platforms, it can also be extremely costly to transfer all of the data from a cloud platform into a data lakehouse, if only due to the cloud platform's (often quite high) egress charges.

Data for Everyone

In addition to the challenges of integrating data, organizations have a broader challenge making data accessible to a wider range of users. Data democratization is providing access to data and analytic resources to everyone at all levels of the organization to empower them to make informed decisions. When you empower employees with data, you enable them to evaluate the processes and procedures more specific to them. This encourages them to identify inefficiencies and challenges within their own jobs and instigate change to make improvements.

You cannot enable data democratization using only data at an organizational level. You need data that is specific to lines of business. You need data that is easy to understand, easy to access, and easy to draw information from. An organizational data warehouse is not likely to contain such granular data and it is also not likely going to be easy to read and understand for an average employee.

While a typical employee can benefit from data collected at the business level, they are more likely to require data closer to their position. For example, they might need production data from the machine that they are assigned; or they could need data about retail sales in their region. You want to provide your employees with data they can use to make decisions on their own, and that data can come from a wide variety of places:

Data from the Organization
: Employees might need access to the business process of the organization. For example, an employee working in a sales role may want to use marketing data to derive insights on how to organize their storefront.

Data from the Line of Business
: The employee generates data as part of their individual task, and often this is the data most relevant to them for analysis. This data is also the most time-sensitive, as it reflects the day to day activities of the employee in his or her job.

Data from Outside the Organization
: Insights can be drawn from beyond the organization itself. Information on weather, suppliers, and more can directly influence business decisions at the employee level.

Employees can benefit from a wide variety of data in their jobs, but how do you make this data available to them? More importantly, how do you structure that data in a way that is easy to understand and utilize? Ideally you want a solution that is self service and requires as little maintenance as possible while also being complete and detailed enough to provide the insights they need.

Overcoming Challenges with Logical Data Management

Demand for data at all levels of your organization combined with larger more varied data sets requires a different approach to data management. Data warehouses lack the detailed data and accessible format needed for democratization. Comparatively, data lakes and data lakehouses demand additional features. Furthermore, the business data environment is complex, leading to poor quality data that makes it difficult to utilize. What options does a company have to meet today's data demands? In the next chapter, we'll introduce you to logical data management, and how it can radically transform a data infrastructure, without having to replace any existing solutions.

CHAPTER 2

Understanding Logical Data Management

A Note for Early Release Readers

With Early Release ebooks, you get books in their earliest form—the author's raw and unedited content as they write—so you can take advantage of these technologies long before the official release of these titles.

This will be the 2nd chapter of the final book.

If you have comments about how we might improve the content and/or examples in this book, or if you notice missing material within this chapter, please reach out to the editor at *vwilson@oreilly.com*.

Now that we've covered some of the challenges with managing data, the imperative is clear: Organizations need a strategy that empowers organizations to wield data with confidence, providing governed, secure, and timely information for business imperatives. Such a strategy must encompass a comprehensive framework, detailing the requisite technology, processes, personnel, and policies to manage an entity's informational assets sustainably. It should bring to life the organization's vision for data acquisition, preservation, sharing, and application. The resulting infrastructure from this strategic blueprint should refine data management at every juncture, ensuring accessibility for all involved parties. The aim is to arm the business

with the data it needs to secure a competitive edge and chart a course for success. This endeavor calls for a deep dive into logical data management, a concept that will be defined and explored in the following section.

Defining Logical Data Management

Logical data management is a transformative approach to data handling within an organization. It moves beyond the conventional constraints of physical data consolidation and instead, adopts a dynamic, virtualized strategy. This approach enhances the integration and accessibility of data, making it a more flexible and efficient resource for the organization.

As the key technology for logical data management, data virtualization creates a common data layer on top of distributed data sources, making it faster and easier to combine data from different sources. It improves accessibility, allowing data consumers and business decision-makers to access the data they need when they need it, expressed in the language of the business, without being restricted by physical limitations. It also fosters agility and data consistency, as the data products delivered by the data layer are based on reusable components representing common business entities. In essence, it is about leveraging technology to make data more flexible, integrated, and accessible, thereby empowering organizations to make more informed decisions. It's a move towards viewing data not just as a static asset, but as a dynamic resource that can drive business growth and innovation.

At its core lies the principle of abstraction, which separates the data layer from the confines of physical storage mechanisms. This separation is crucial, as it grants businesses the agility to access data from a multitude of data sources, irrespective of their location—be it cloud-based systems, on-premises databases, or a fusion of both. It also allows business data needs to evolve independently of the changes in the underlying data infrastructure provoked by cloud transition and IT modernization initiatives. The implementation of a logical data layer is akin to removing barriers within an organization, dismantling the data silos that have long hindered a unified view of the information landscape and fostering a collaborative environment where data managers and consumers can work in concert. A logical data layer can be implemented above any existing

data infrastructure, including cloud data warehouses and data lakehouses.

The strategic value of logical data management is significant. It propels organizations into a future of rapid decision-making, efficient operations, constant innovation, and prioritized customer satisfaction. For example, a multinational corporation can leverage it to integrate customer data from various touchpoints, enhancing service delivery and marketing strategies. Similarly, financial institutions can simplify governance and stay ahead of regulatory changes by centralizing policy management in the common logical data layer for all data sources and consumers. As you continue into the next section, you'll learn about more benefits that logical data management brings to the table.

The Benefits of Logical Data Management

As a leader in your organization, your days are filled with critical decisions that hinge on the accuracy and availability of data. Traditional data management systems might leave you waiting for IT to consolidate reports from various departments, a process fraught with delays and potential inaccuracies. Logical data management changes this narrative by enabling a streamlined, real-time view of your company's data landscape that is not conditioned by data infrastructure and guarantees data consistency by design.

It gives you a bird's-eye view of operations, finance, customer interactions, and market trends, all through a single pane of glass. This unified view is not just a convenience; it is a game-changer. For instance, when evaluating the performance of a new product line, logical data management enables you to instantly pull together sales figures, production costs, and customer feedback, painting a complete picture of the product's journey from conception to market reception.

Here are the key benefits of logical data management, with each benefit illustrated by an example so you can envision how it might apply to your own context:

Harmonization of Data Assets
: Enables swift location and utilization of information, allowing a customer service representative to access a unified view of customer data from CRM systems, sales platforms, and support

channels, leading to more efficient and personalized customer interactions.

Enhanced Security and Compliance
: Provides a common data layer to enforce uniform policies across data platforms, ensuring data protection and compliance with regulations like GDPR or SOX, thereby instilling trust in healthcare professionals regarding the protection of patient records.

Fostering Collaboration
: By exposing data in the language of the business, it bridges the gap between data managers and users, allowing marketing teams to work seamlessly with IT to refine customer segmentation models, thereby boosting the effectiveness of marketing campaigns.

Process Improvement
: The common data layer reveals opportunities for process improvement and promotes data reuse and sharing, enabling a supply chain manager on an e-commerce platform to use insights from it to optimize inventory management systems.

Comprehensive Data Perspectives
: Provides comprehensive data perspectives by integrating data holistically, allowing an operations manager in an energy company to merge sensor data with operational and market data for a unified predictive maintenance view.

Operational Efficiency
: Boosts operational efficiency by optimizing data management tasks and automating processes, such as route optimization for logistics managers, saving time and resources.

Improved Data Accessibility and Literacy
: Improves data accessibility and literacy, fostering transparency and civic engagement, and enabling city planners to make public data more accessible to citizens.

Driving Innovation
: Drives innovation by empowering the business with the data they need in formats they can use, and by feeding AI and machine learning algorithms with diverse, quality data, unlocking new possibilities and competitive advantages, including tai-

loring customer interactions in real-time and adapting quickly to market trends.

The Benefits: An Architectural View

A logical data layer reads data from a wide variety of sources and adds transformations that make it appear as if it were a single source. Logical layers are not restricted by the confines of the physical layer formats. You can connect to diverse data types and combine disparate data sources together to create relationships that otherwise would not be possible at a physical level (Figure 2-1).

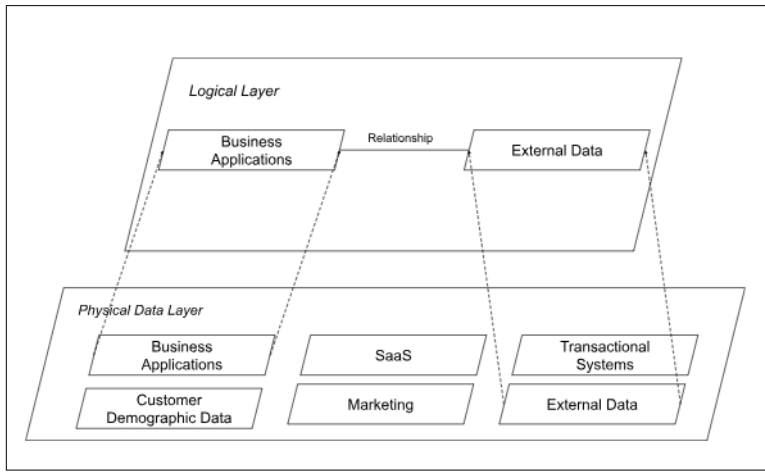


Figure 2-1. A logical layer sits atop the physical data sources and provides a lens that translates the data. This allows disparate data sources to be combined with user defined relationships without having to always replicate data.

The data itself still resides in its original location, meaning it does not always need to be replicated as part of the process (It *can* be, but it need not be). You can think of a logical data layer as a library, where you have access to the knowledge it contains without having to obtain your own copy of every book. The logical layer pulls only the necessary fields from the variety of physical data sources and uses its own language to create new fields, tables, and relationships between the different data sources. Since the data does not need to be physically copied, the amount of resources required to maintain the layer is much smaller.

The logical layer simplifies life for your data analysts, as it is located centrally under a single format. Your analysts no longer need to have knowledge of all the different database languages your company uses. Data stored in SQL, Oracle, or any other format is translated into the logical layer, which means your analysts only need to be familiar with the logical layer's language.

Additionally, the logical layer is not restricted to the nomenclature of the physical layer. This means your developers can rename tables and fields to common names used throughout the business. This makes it much easier to use and accessible to your employees with less experience dealing with the data tables. It also provides a way for your business to connect to process-specific data sets, enabling your less technical employees to utilize data to make decisions much closer to their line of work.

The data is also easily reusable. A business that utilizes logical data management can set up tables with key information related to the core functions of the business. These in turn can be used to create domain specific data products. For example, a business can generate data specific to products. This can then be used to generate reports using consistent data tailored to business domains such as by marketing, sales, or regional locations.

The logical layer is extremely flexible, and makes it easy to add additional data sets. Since the layer sits on top of the data, there is no need to copy data from one location to another when a new system is added. The logical layer can easily be configured to point at the new source and extract the fields needed for reporting purposes. This includes process specific data sets, such as machine usage, point of sales, and CRMs.

Since there is no need to copy data, access to data can be closer to real time. For example, if your business has monitoring software that tracks production, you want that data to be as close to real time as possible. A logical layer can sit atop the production data and expose it for use without the need to copy it to a reporting database. This ensures that the employees responsible for maintaining that production line have access to that data and can make faster decisions when issues arise.

As new source systems arise and current systems update, the tables and fields reflected in the logical layer do not need to change. The structure of the logical layer adapts to these changes, keeping a

consistent and familiar view for your users. This also reduces the need to recreate or modify reports as the systems and underlying sources are modified. This is especially true when transitioning your business to the cloud.

For your users, a logical data layer improves visibility into data lineage. One question that often arises with reporting is "where did this data come from?" For traditional data management systems, this question can sometimes be challenging to answer. When the data is pulled from the production systems, transformed, and loaded elsewhere the source field and tables can get lost. Tracing a field in a report back through the reporting and ETL changes can be anywhere from challenging to impossible for end users. A logical layer simplifies this by simply providing a lens on the data instead of copying it.

This clear view into how the data moves also facilitates better data definitions, data governance, and simplifies security. Instead of managing security for multiple data sources, administrators can focus on access to the logical layer. This removes the need for access expertise and management in multiple systems, requiring the administrators to focus instead on the single logical layer access.

Logical data management transcends traditional data management, propelling organizations into a future-ready, data-empowered state. By understanding and leveraging its strategic benefits, you can harness data more effectively, drive innovation and secure a competitive edge in the marketplace. The next section will uncover how logical data management can enhance various leadership roles.

Empowering Leadership Roles with Logical Data Management

Let's explore how logical data management can empower you to make better decisions by looking at a few more examples through the lens of various leadership roles.

IT Leaders: Navigating the Digital Maze

: According to the 2024 Gartner CIO and Technology Executive Survey, budget constraints loom large, making innovation a delicate balance with efficiency. Consider an Infrastructure Leader responsible for modernizing the organization's IT infrastructure. Logical data management becomes their compass:

they integrate data on infrastructure costs, performance metrics, and resource utilization. By identifying inefficiencies and reallocating resources strategically, they optimize the technology budget. With seamless data access across legacy systems, cloud platforms, and emerging technologies they can make data-driven decisions during digital transformation initiatives. For instance, they might enable real-time analytics for customer insights or streamline DevOps processes. Moreover, with logical data management, the Infrastructure Leader unifies data silos, integrates on-premises and cloud environments, and enables agility (e.g., they might orchestrate hybrid solutions, ensuring smooth migration and scalability). Even within budget constraints, they're able to facilitate efficient, innovative solutions.

Business Leaders: Crafting Exceptional Customer Experiences
: Business Leaders hold the key to shaping exceptional customer experiences. They understand that timely data is critical for informed decision-making, yet they often face a dilemma: reliance on IT teams for data access. Imagine a Chief Marketing Officer (CMO) aiming to personalize marketing campaigns. With logical data management, the CMO gains direct access to customer insights, behavioral patterns, and preferences. By analyzing data across all relevant data sources, the CMO tailors marketing messages, predicts trends, and delivers personalized experiences. Similarly, a Chief Operations Officer (COO) can optimize supply chains, reduce lead times, and enhance product availability by accessing integrated real-time data across warehouses, suppliers, and logistics partners. By breaking free from IT bottlenecks, business leaders can stay agile, exceed customer expectations, and drive innovation.

Data Leaders: Governing the Data Estate
: Data Leaders are the custodians of the data kingdom. Their mission is to simplify data governance while promoting a data culture that enables self-service capabilities. Logical data management can help them achieve this, ensuring data quality, integrity, and compliance with regulations. The advent of low-code and no-code connectivity solutions, coupled with a reduced reliance on Extract, Transform, Load (ETL) processes means there is less need for many developers dedicated solely to data management. Instead, existing developers can focus on delivering value-added solutions and products to the business at

a faster pace, without the need to constantly build complex ETL processes or support monolithic architectures.

Compliance Leaders: Shielding Against Risks

Compliance Leaders are the guardians of the organization, protecting it against evolving risks and compliance requirements. Logical data management can equip them with the agility to respond to new types of risks or compliance regulations as they arise. For data consumers, this means they can trust the data they are using. Logical data management ensures that data is accessed in a manner that respects individual privacy rights and adheres to ethical standards. This includes mechanisms for consent management, data anonymization, and secure data handling practices. It provides a framework to ensure compliance with various data protection regulations and ethical guidelines. This not only mitigates legal risks but also enhances the organization's reputation for responsible data management. Moreover, a well-governed logical data management strategy promotes transparency in data processes. It allows both data consumers and decision-makers to understand how data is being used, fostering trust and ethical use of data within the organization.

Operations Leaders: Turbocharging Operational Excellence

Operations Leaders, the driving force of an organization, are under constant pressure to enhance efficiency, productivity, and quality in the realms of Manufacturing & Supply Chain, Financial Operations (FinOps), Marketing Operations (MarOps), Infrastructure Operations and more. Logical data management can enable real-time, data-driven decision-making for improved operational performance. Gartner reveals that 85% of infrastructure and operations leaders aim for more automation in the next two to three years.[1] Deloitte's study indicates that 86% of manufacturing executives see smart factory solutions as primary competitiveness drivers in the next five years.[2] However,

1 Gartner, "Gartner Survey Finds 85% of Infrastructure and Operations Leaders Without Full Automation Expect to Increase Automation Within Three Years", *https://www.gartner.com/en/newsroom/press-releases/2022-10-03-gartner-survey-finds-85-percent-of-infrastructure-and-operations-leaders-without-full-automation-expect-to-increase-automation-within-three-years*, October 3, 2022

2 Deloitte, "2024 manufacturing industry outlook", *https://www2.deloitte.com/us/en/insights/industry/manufacturing/manufacturing-industry-outlook.html*, Oct 30, 2023

a Forrester report shows only 12% of infrastructure and operations leaders exceed CIO expectations, suggesting a significant scope for improvement through effective data management.[3]

Financial Leaders: Pioneering the Digital Frontier
: Financial Leaders play a pivotal role as navigators, guiding organizations through complexities. Logical data management serves as their compass, addressing critical challenges. It optimizes costs by eliminating data redundancy and streamlining resource allocation. It ensures compliance with regulations, safeguarding sensitive financial information. By providing agile access to integrated data, logical data management empowers timely decision-making, enabling strategic initiatives to thrive within budgetary constraints.

Summary

The future of logical data management is promising, with trends indicating a continued shift towards distributed data management. As organizations deal with data and applications that remain distributed across regional and cloud boundaries, logical data management will play a crucial role in managing this distributed data efficiently and cost-effectively.

For data consumers, logical data management acts as a bridge, seamlessly connecting disparate data sources while ensuring security and compliance. Data discovery becomes intuitive, and the power of information lies just a query away.

Logical data management offers substantial benefits, including improved decision-making and data consistency. However, organizations should consider challenges such as implementation effort, initial costs, resource allocation, and balancing rigor with agility. Smaller companies may need to weigh the benefits against their investment. Legacy systems, data governance, scalability, and complexity also play a role. A thoughtful approach and ongoing evaluation are crucial for its successful adoption.

---

3 Gartner, "Gartner Survey Finds Only 12% of Infrastructure & Operations Leaders Exceed CIO Performance Expectations", *https://www.gartner.com/en/newsroom/press-releases/2023-12-6-gartner-survey-finds-only-12-percent-of-infrastructure-and-operations-leaders-exceed-cio-performance-expectations*, December 6, 2023

In the next and following chapters, you'll see how logical data management is being put into action today, to solve intricate business and technical challenges with efficiency and intelligence.

About the Author

Christopher Gardner is a business intelligence analyst and lead Tableau developer for the University of Michigan. He has over 24 years of experience with the university and has served as a data analyst and Tableau expert for the campus since 2013. Within his role for the university and through O'Reilly boot camps and classes, he has taught thousands of users how to develop data visualizations and dashboards within Tableau. In addition, Christopher is a tech editor and writer for O'Reilly, participating in various Tableau and data-related articles and books. He is a Tableau Certified Data Analyst and has maintained an equivalent of that since 2016. He holds a degree in Actuarial Mathematics from the University of Michigan.