

An Introduction To Data Mesh: What It Is And How It Works



Table Of Contents



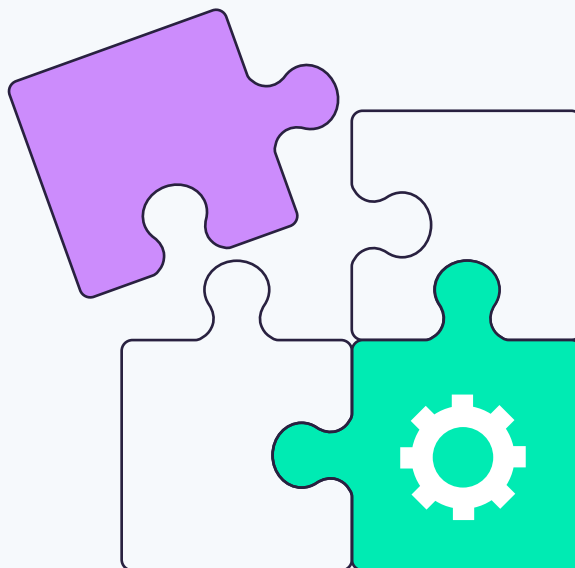
INTRODUCTION	3
WHAT IS A DATA MESH?	4
HOW DOES A DATA MESH WORK?	5
WHAT ARE THE FOUR KEY PILLARS OF A DATA MESH?	7
HOW IS A DATA MESH DIFFERENT FROM TRADITIONAL DATA MANAGEMENT ARCHITECTURES?	12
WHY DO YOU NEED A DATA MESH?	14
BENEFITS OF A DATA MESH ARCHITECTURE	15
WHAT ARE THE CHALLENGES OF IMPLEMENTING A DATA MESH?	17
WHAT ARE SOME OF THE BEST PRACTICES FOR DATA MESH DESIGN AND IMPLEMENTATION?	18
IS DATA MESH RIGHT FOR YOUR DATA ECOSYSTEM?	20

01

INTRODUCTION

With the rising popularity of big data, there is a growing need for ways to manage and analyze this data. Data mesh is a fresh approach to data management that has recently grown popular. It is designed to address the challenges of managing large, complex data sets in agile and scalable ways to respond effectively to changing business needs.

In this whitepaper, we will introduce the concept of data mesh and discuss how it works to manage big data. We will explore how it varies from traditional data management techniques and approaches. We will also discuss the key pillars and advantages of a data mesh architecture. On top of it all, we will provide some best practices of how your organization can effectively implement a data mesh framework.



02

WHAT IS A DATA MESH?

A data mesh is an architectural design for data management and access defined by Zhamak Dehghani. It is designed to provide decentralized access to manage and share [big data](#) at the domain level.

The traditional data management approach involved a centrally managed data ecosystem in the form of data fabric, warehouses, and lakes. A data mesh is designed to address some of the challenges that can arise in traditional data management systems, such as a lack of clear ownership or accountability, difficulty accessing and using data, and difficulty maintaining a consistent and accurate view of the data. A data mesh helps by simplifying the flow from data production to data consumption.

The idea behind a data mesh is to have each domain team process, manage and publish relevant data via a self-serve platform that

organizations can use across business domains for various purposes.

In a data mesh architecture, data is treated as a first-class citizen within the organization, with clear ownership and governance structures in place to ensure that it is used consistently and effectively. Data is organized into logical domains, and domain experts are responsible for the quality, consistency, and usability of the data within their domains. Data mesh also emphasizes using standardized, low-ceremony interfaces for accessing and using data. This helps minimize the barriers to accessing data by removing the bottleneck of central IT control over data access, making it easier for people across the organization to use data effectively.

A data mesh aims to remove the common bottlenecks that organizations face with their data pipelines and lifecycle to enable a more scalable, flexible, and decentralized approach to data governance and usage within an organization.

How Does a Data Mesh Work?

A distributed data mesh is a form of a microservices architecture. However, a data mesh follows a domain-driven design instead of coupled services. A data mesh sets parameters around data using functional domains, enabling data to be treated as a product that domain leaders across the organization have decentralized access. A data mesh architecture typically involves the following components:

1 BUSINESS DOMAINS

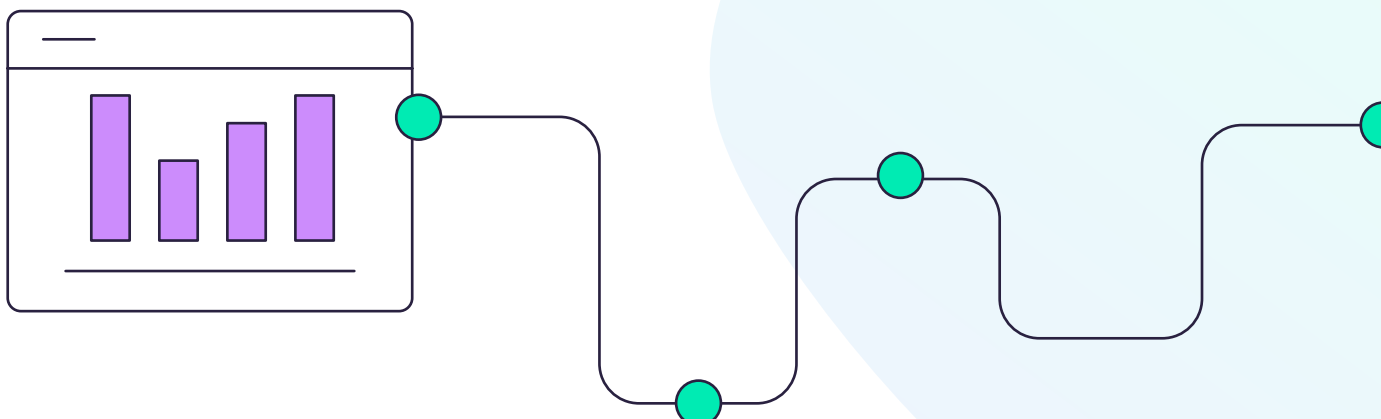
Data is organized into logical territories, each corresponding to a specific business unit or a particular type of data. Domain experts are responsible for the quality, consistency, and usability of the data within their domains.

3 DATA CONSUMERS

These are the people that consume data, particularly for analytics. It also includes all systems or processes that use data from the data mesh.

2 DATA PRODUCERS

These are the systems or processes that generate data and make it available for consumption by other parts of the organization. Data producers also include subject matter experts who document their domain data's semantic definitions, catalog the metadata, and set policies for its consumption.



1 DATA MESH PLATFORM

This is the infrastructure that enables the flow of data between data producers and data consumers and manages the governance and access controls for the data.

2 DATA LINEAGE AND TRACEABILITY

Data mesh architectures typically include mechanisms for tracking data flow through the system and understanding how it has been transformed or modified over time.

The data mesh has evolved the enterprise data culture by treating data as a product, instead of a by-product, of a process. Under a data mesh model, each domain team is responsible for managing its [ETL data pipelines](#). Since each domain team can have its own unique use cases, the data mesh helps by ensuring that each team has access to the relevant analytical data to perform tailored analysis.

These domain teams are subject matter experts who can leverage their domain's data to design APIs in their data consumers' best interests. However, domain teams are not a replacement for central data engineering teams. Data engineers in a data mesh have a critical role in determining the best data infrastructure solutions for the data products from every organizational domain.

To use a data mesh, a data consumer typically starts by identifying the data they need and the domain it belongs to. They can then use the data mesh platform to access and use the data, potentially transforming or enriching it as required. The data mesh platform ensures that the data is governed and managed according to the established policies and procedures for the relevant domain.

A data mesh architecture provides flexible and interoperable integrations that work with decentralized data governance structures. This enables organizations to better leverage their data products for immediate data consumption from multiple domains. As a result, they can effectively reach their data goals for analytics, data science experimentation, and other data needs.

03

WHAT ARE THE FOUR KEY PILLARS OF A DATA MESH?

A data mesh architecture rests on four key pillars that regulate the main objectives of the mesh model. Before getting into the pillars of a data mesh, it is important to understand the key objectives behind the data mesh architecture:

1 HIGH QUALITY

The data mesh model should provide high data quality. This is contrary to the high-quantity goal of a centralized data architecture.

2 HIGH INTEROPERABILITY

A data mesh should achieve high interoperability for the smooth utilization of data products between teams.

3 DATA DISCOVERABILITY

Data products in a data mesh should be promptly discoverable to relevant data teams as and whenever required.

4 DATA SECURITY

Data products can be difficult to protect in a decentralized data mesh. Therefore strict security guidelines should be implemented and monitored to maintain high security in the mesh.

With the above-mentioned objectives in mind, here are the four key pillars of an effective data mesh architecture:

- Data-As-A-Product
- Federated Governance
- Self-Service Infrastructure
- Domain-Oriented Ownership

Data-As-A-Product

To ensure interoperability across different business domains, a data mesh takes a data-driven approach by treating data as a product of business processes. Data-as-a-product refers to the idea that business teams treat data as a valuable asset that can be developed, managed, and monetized similarly to traditional products. This approach to data management involves treating data as a strategic resource that can drive business value and generate competitive advantage.

In a data mesh, data is taken as a product owned and managed by a cross-functional team of data experts rather than a centralized IT department, data governance team, or data engineers. Consequently, data engineering teams can focus on implementing discoverable, explorable, secure, and reliable infrastructure that works across domains to make it easier for domain experts to create and manage their data products.

This decentralized approach to data management allows for more agile and flexible data governance, as well as greater ownership and accountability for the quality and value of the data.

Data-as-a-product needs a shift in mindset and approach to data management, as it requires organizations to prioritize the development and management of high-quality data assets. It includes investing in the necessary tools and resources to ensure the accuracy, completeness, quality, and accessibility of data, as well as establishing clear roles and responsibilities for data ownership and stewardship.

For example, the sales team can create a clear sales data set that can easily be utilized by a data science team for some machine learning model.

Self-Service Infrastructure

Self-service infrastructure is a critical concept in data mesh architecture. It refers to the ability of teams and individuals within an organization to access and use data without relying on IT. The self-serve practice empowers teams to have more control and autonomy over their data. It enables them to be more agile and responsive in their use of data.

In a data mesh architecture, self-service infrastructure is achieved through standardized data interfaces, data pipelines, and data catalogs. These tools allow teams to discover and access the needed data without requiring specialized technical expertise or a lengthy request process.

To enable self-service infrastructure, organizations must also establish clear policies and guidelines for data use and governance and provide training and support to help teams understand how to use data effectively and responsibly.

For instance, the finance team might use some financial data set previously created by a data scientist. The data set should also include precise metadata details that can quickly tell the finance team what each column in the data set means. These details include data semantics, creation dates, usage, etc.

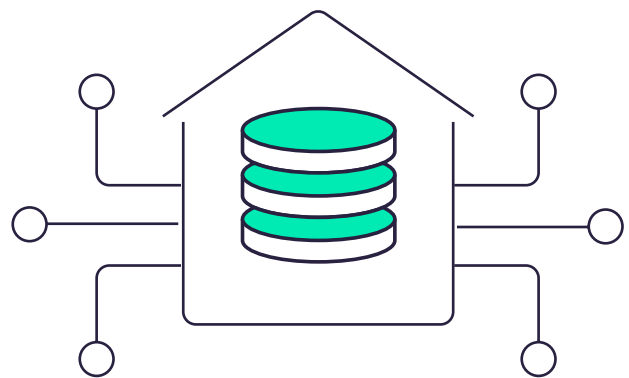
Federated Governance

The decentralized architecture of a data mesh makes it crucial to create and follow standards to ensure consistency across data products and prevent data duplication. So a data mesh implementation usually involves a central team that outlines certain practices that need to be followed by all domain teams when publishing their data products.

Federated governance is the decentralized approach to [data governance](#) central to the data mesh framework. In a data mesh architecture, data governance is not only the responsibility of a centralized IT department or data governance team. It is the shared responsibility of all teams and individuals who use and produce data. This approach to data governance is achieved through federated governance standards.

Federated governance standards allow teams more control and autonomy over their data while adhering to overall organizational policies and all data governance guidelines. In a data mesh architecture, federated

governance structures typically include data stewards, who ensure the quality and integrity of data within their domain. Moreover, there are data councils, which are cross-functional groups that oversee data governance at a higher level and provide guidance and support to data stewards.



An example of implementing federated governance can be setting file formats and naming conventions that need to be used when creating a data set.

Domain-Oriented Ownership

Domain-oriented ownership refers to the idea that teams and individuals within an organization should be responsible for the data relevant to their domain of expertise.

This approach to data ownership is designed to ensure that data is appropriately managed, governed, and used to drive value within specific business areas. By maintaining data sets closer to their domain experts, domain-oriented ownership helps to accelerate data processes and ensures richer data sets.

A data mesh organizes data into logical domains corresponding to specific business areas and data types. Domain experts are responsible for the data quality, consistency, and usability within their disciplines. Therefore, the functional data unit in a domain-driven approach is not a pipeline stage. Instead, the entire domain is a valuable asset that manages, uses, and serves this data. Data independent of a central unit of experts and infrastructure

creates a culture of ownership, which was a significant bottleneck of a central data lake where it was difficult to identify domain owners of a certain dataset.

In a data mesh architecture, domain-oriented ownership is achieved through data stewards, who ensure the quality and integrity of data within their domain of expertise. Data stewards are domain experts who deeply understand the data and its value to the business and are responsible for managing and governing the data within their domain.

Each team is responsible for ensuring the operability of its data pipeline, which enhances the quality of the data assets of the entire organization. Data products created and owned by the domain teams make it easy for everyone to identify the domain owners. Communication amongst different groups is also significantly streamlined, as teams can approach product owners without any hassle.

04

HOW IS A DATA MESH DIFFERENT FROM TRADITIONAL DATA MANAGEMENT ARCHITECTURES?

Data mesh is a relatively new architectural pattern for data management and access distinct from other approaches, such as the traditional data lakes and warehouses. Here are some key ways in which data mesh differs from traditional data management architectures:

Self-Service Infrastructure

Data lakes are large, centralized repositories that store big and small data in structured or unstructured forms. On the other hand, data mesh architectures are more decentralized and store data in different domains, with domain experts responsible for maintaining the data within their domains.

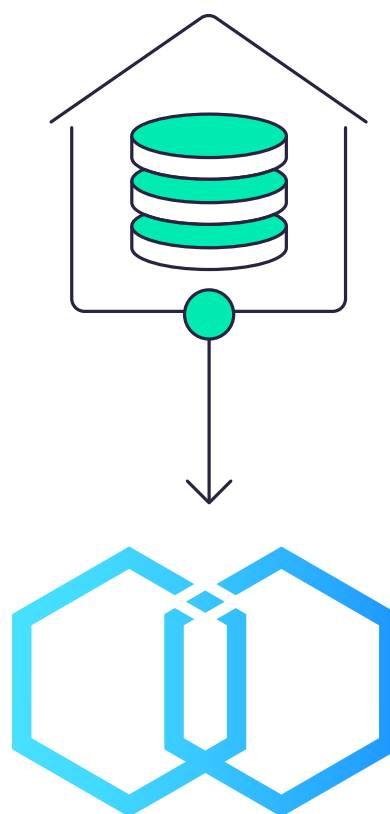


Data Mesh vs. Data Warehouses

Data warehouses are centralized systems that store structured data extracted from various sources and transformed for analysis. Data mesh architectures are more decentralized and do not necessarily require data to be structured or transformed in a specific way before its usage.

Whether we talk about data lakes or warehouses, there are several problems that all centralized systems have in common. Firstly, with a central data engineering team managing access, it can lead to high latency if more and more access requests start coming from the domain teams. Secondly, in a centralized system, pre-processing a data source is the responsibility of the engineering team. However, this requires much domain-specific knowledge, which a single data engineering team may have yet to have. Moreover, it isn't easy to highlight systems or processes that are the actual data producers in a centralized system. It is just one single team preparing and delivering a data set.

Basically, a data mesh moves towards a more democratized system where each domain team can manage its data. In comparison, a centralized repository is managed by a single data engineering team that handles all the access and delivery issues across the entire organization.



05

WHY DO YOU NEED A DATA MESH?

There are several reasons why you might choose to implement a data mesh architecture. Organizations typically transition to the data mesh architecture when data ownership is centralized, creating a big monolithic data platform. A monolithic data architecture leads to data silos where it becomes difficult for data platform engineers to ensure access to the correct data. These engineers have no understanding of the business domain or use cases.

A decentralized approach to data management enables organizations to include clear governance structures that ensure data is used consistently and effectively across your organization.

You can use data mesh to promote data literacy across your business by providing standardized interfaces for accessing and using data and by providing tools and training to help people understand how to work with data effectively. Moreover, you can track the data flow through the mesh and know how it has been transformed or modified over time.

Another crucial reason for opting for a data mesh is the decentralized responsibility for data quality and usage. Domain experts and data consumers have clear ownership and accountability for their data.

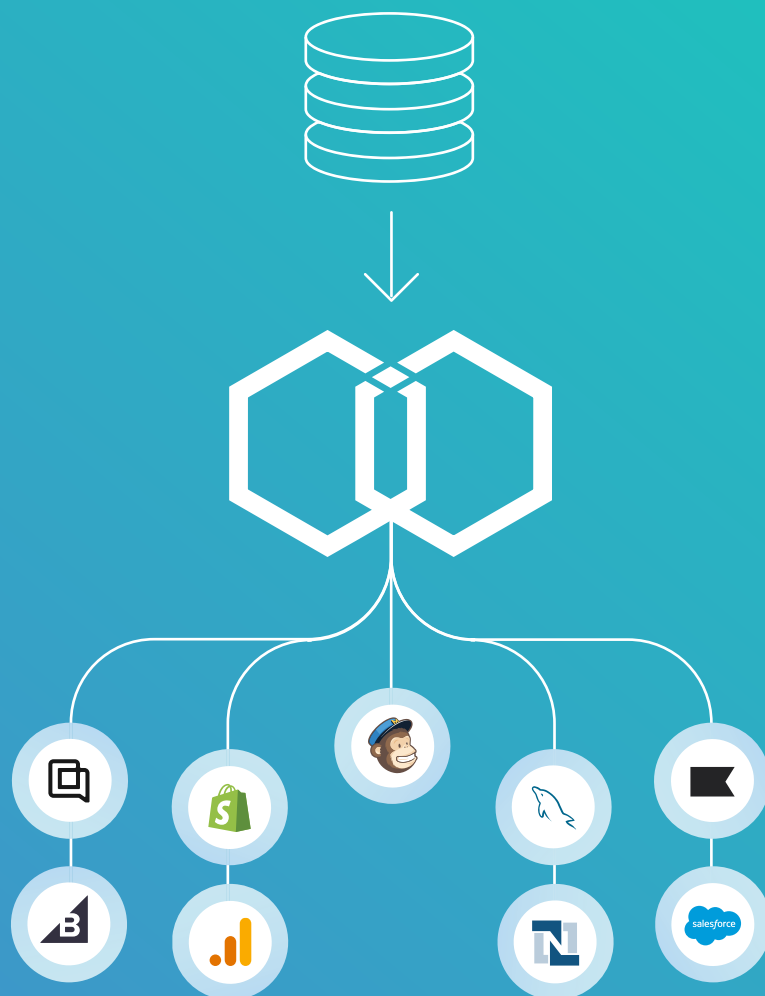
And most importantly, data mesh architectures are designed to be adaptable and flexible, allowing them to evolve as the needs of your organization change.

06

BENEFITS OF A DATA MESH ARCHITECTURE

Following are some of the many benefits of implementing a data mesh architecture:

- Data Democratization
- Improved Data Quality
- Scalability
- Agility
- Data Security
- Better Decision-Making
- Innovation



DATA DEMOCRATIZATION

Datasets in a data mesh are readily accessible by enterprise-wide users. Users do not require domain-specific expertise to work with these datasets, resulting in effectively utilized datasets.

IMPROVED DATA QUALITY

With every data domain assigned to specific domain experts, data mesh ensures high data quality within each domain, resulting in rich data products.

SCALABILITY

Data mesh enables decentralized data operations and independent teamwork to provide a scalable data ecosystem with reduced operational and storage costs.

AGILITY

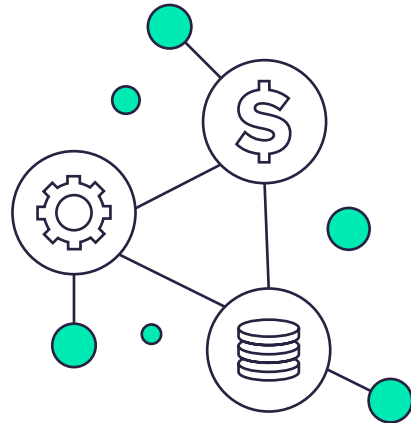
The adaptable capabilities of a data mesh increase an organization's agility to respond more quickly to changing business needs.

DATA SECURITY

Instead of circulating data copies across public networks, data mesh asks for data where it is already stored, eliminating the risk of data leakage and loss.

BETTER DECISION-MAKING

By making data easily accessible and usable for people across the organization, a data mesh architecture assists teams in reaching meaningful conclusions.



INNOVATION

Data mesh architecture empowers innovation by enabling organizations to take a giant step forward by moving from complex and centralized to simple, decentralized data.

07

WHAT ARE THE CHALLENGES OF IMPLEMENTING A DATA MESH?

While a data mesh is an innovative step in data management, it also comes with challenges. Implementing a data mesh architecture can be challenging for several reasons:

ENSURING EFFECTIVE DATA MANAGEMENT PRACTICES

High-quality data within a data mesh architecture can be challenging, particularly if the organization has a history of poor data management practices. Effective implementation of a data mesh architecture requires careful planning and a commitment to changing data management and usage within the organization.

ENFORCING STRICT DATA GOVERNANCE

Establishing decentralized data governance structures within a data mesh architecture can be challenging, particularly in organizations with centralized data governance. It requires the involvement of all key stakeholders to define roles and responsibilities and establish clear policies and procedures for data usage.

CHANGING THE EMPLOYEE MINDSET

In an organization where employees are used to the traditional centralized management techniques, it can be challenging to introduce and implement a data mesh approach. It requires a complete mindset change about data processing using a concerted effort to educate and train people across the organization on the benefits and best practices of a data mesh approach.

TECHNOLOGICAL CHALLENGES

It can be difficult to choose the right technologies and implement an appropriate tool suite to ensure a competent data mesh. Failing to set up the right technologies can lead to challenges in implementing a data mesh.

08

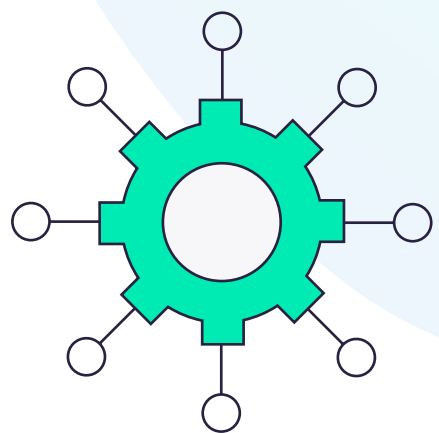
WHAT ARE SOME OF THE BEST PRACTICES FOR DATA MESH DESIGN AND IMPLEMENTATION?

To avoid the challenges of setting up a data mesh, here are some best practices for data mesh design and implementation:

- Start small and scale incrementally
- Engage domain experts
- Establish clear governance structures
- Invest in data literacy
- Plan for data integration
- Monitor and optimize performance

START SMALL AND SCALE INCREMENTALLY

It can be tempting to try to implement a data mesh architecture across the entire organization all at once, but this can be overwhelming and may not be practical. Instead, consider starting with a small project focusing on a specific area or use case. Then, scale incrementally as you learn and iterate.



ENGAGE DOMAIN EXPERTS

A data mesh architecture heavily relies on domain experts to ensure data integrity and accuracy in their domains. It is essential to give domain experts ownership over the data and engage them early in the design process and involve them in developing data policies and procedures.

ESTABLISH CLEAR GOVERNANCE STRUCTURES

Data mesh architectures require clear and decentralized governance structures to ensure data is used consistently and effectively across the organization. Defining roles and responsibilities and transparent policies and procedures for data usage are vital for a successful data mesh implementation.

INVEST IN DATA LITERACY

Data mesh architectures aim to promote data literacy across the organization, but this requires investment in training and education. Consider providing resources and support to help people understand how to work with data effectively.

PLAN FOR DATA INTEGRATION

A data mesh architecture typically involves the integration of data from multiple sources and systems. It is essential to plan for data integration upfront and consider how data will be transformed and enriched as it flows through the system.

MONITOR AND OPTIMIZE PERFORMANCE

Data mesh architectures can generate large volumes of data. It is crucial to monitor and optimize performance to ensure the system runs efficiently. Consider using tools and techniques such as data lineage and traceability to understand how data flows through the system and identify potential bottlenecks or issues.



09

IS DATA MESH RIGHT FOR YOUR DATA ECOSYSTEM?

A data mesh is indeed the future of data management. However, organizations need to rethink the architectures they have in place and then consider whether a change is required.

A data mesh is only for some. Suppose you are a small organization, not facing any problems with the central data team. In that case, you are better off with a centralized data warehouse. However, suppose your organization works with big data, experiences rapid expansions, and has to deal with a lot of different data sources coming in. In that case, a data mesh might be the way to go.

While contemplating implementing a data mesh, you can make the best of your existing data warehouses with [Integrate.io](https://integrate.io). With Integrate's ETL platform, you can monitor your data warehouses, jobs, and crucial dependencies between your queries and tables to take timely action. Try it for yourself — [sign up for a 14-day trial](#) and [book a consultation](#) with us so we can talk about your data management needs.

We hope that you have enjoyed reading this whitepaper.

We hope that you have enjoyed reading this whitepaper. If so, be sure to check out our other Whitepapers related to ["Data Mesh: What It Is And How It Works"](#).

Learn more about how Integrate.io can help

Make your data work for you!

Chat with an Expert!

