

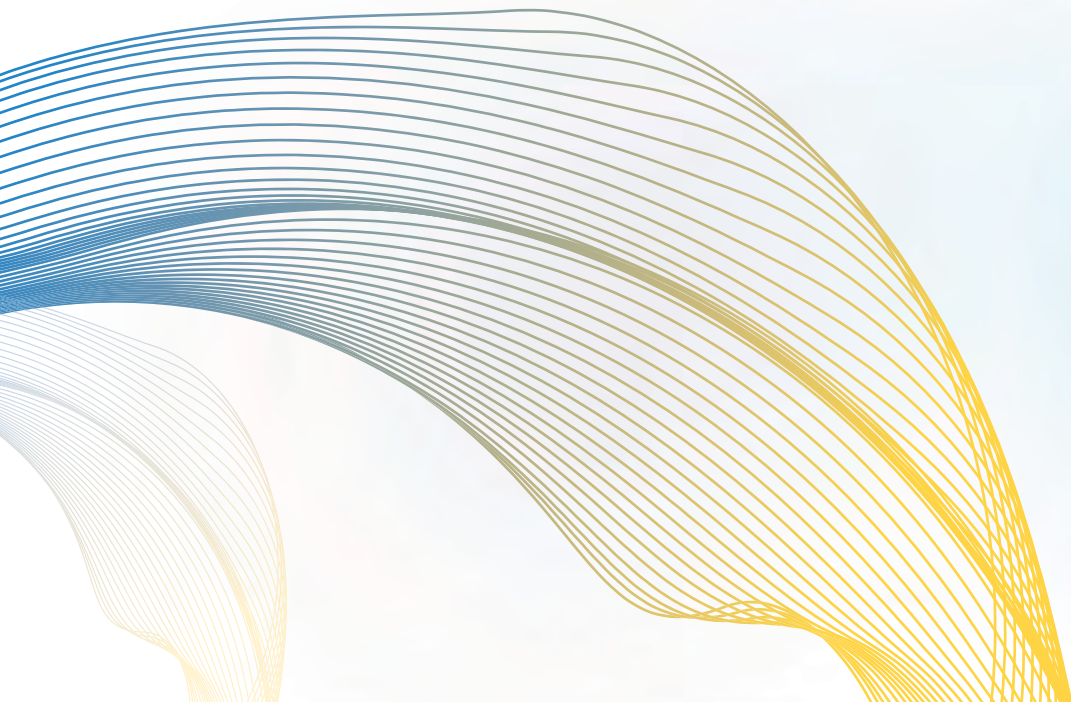
Point of View



Data Products Using **Data Mesh** for Different Industries

Authored by:

Dharmender Verma



Abstract

A short horizontal bar with a gradient from blue to yellow.

This paper is a quick intro to data mesh and how it's different from traditional centralized data management methods like data warehouses and data lakes.

It also talks about the four core pillars of data mesh along with its benefits, challenges, use cases, and industry examples.

Data Mesh

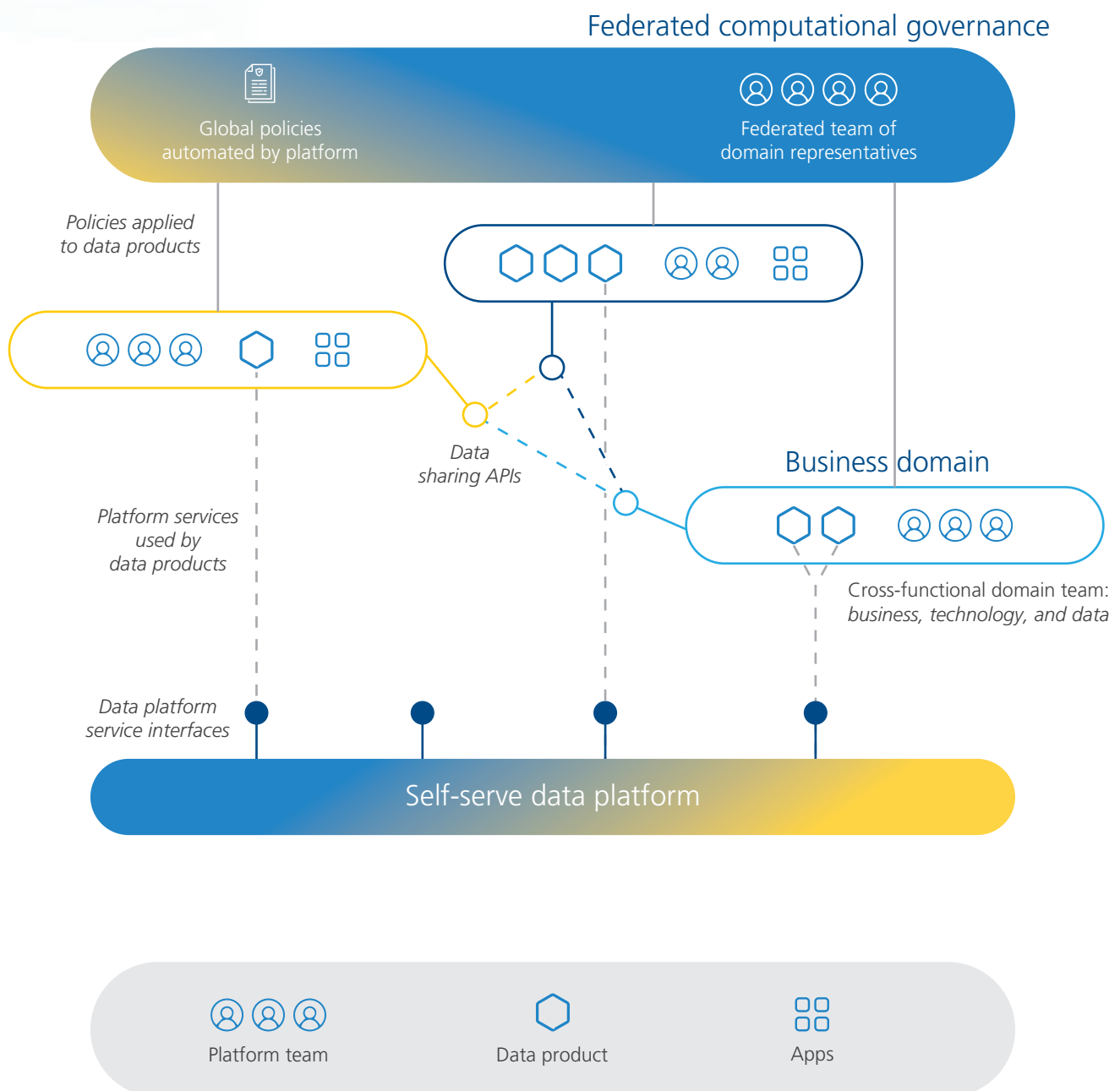
Data mesh is a strategic approach to modern data management and a way to strengthen an organization's digital transformation journey, as it centers on serving up valuable and secure data products. The main objective of a data mesh is to evolve beyond the traditional centralized data management methods of utilizing data warehouses and data lakes. Data mesh emphasizes the idea of organizational agility by empowering data producers and consumers to access and manage data, without the trouble of delegating to the data lake or data warehouse teams. The decentralized method of data mesh allocates data ownership to domain-specific groups that serve, own, and manage data as a product.



Data mesh introduces multidimensional technical and organizational shifts from earlier analytical data management approaches. The below diagram summarizes the shifts that data mesh introduces, compared to previous approaches.



Data Mesh at a Glance



Domain-driven data ownership and architecture

To understand domain-driven data, we must know what a domain is. A domain is an aggregation of people organized around a common functional business purpose. Data mesh makes a domain responsible for the management of data related to and created by the business function of the domain. The domains are responsible for the assimilation, transformation, and provision of data to the end users. Eventually, the domain exposes its data as data products, whose entire lifecycle is owned by it.

Data as a product

Data products are produced by the domain and consumed by downstream domains or users to create business value. Data products are different from traditional data marts, as they are self-contained and are responsible for aspects such as security, provenance, and infrastructure concerns related to ensuring that the data is kept up to date. Data products enable a clear line of ownership and responsibility and can be consumed by other data products or by end consumers directly, to support Business Intelligence (BI) and Machine Learning (ML) activities.

Self-serve data platform

The concept of self-serve data infrastructure is that it is made up of numerous capabilities that can be easily used by members of the domains to create and manage their data products. The self-serve data platform is supported by an infrastructure engineering team, whose primary concern is the management and operation of the various technologies in use. This illustrates the separation of concerns. Domains are concerned with data, and the self-serve data platform team is concerned with technology. The measure of success of the self-serve data platform is the autonomy of the domains.

Federated computational governance

Traditional data governance can be seen as an inhibitor to producing value through data. Data mesh enables a different approach by embedding governance concerns into the domain workflows. There are numerous aspects to data governance, however, when considering data mesh, it is imperative that usage metrics and reporting become part of this definition. The amount of data usage and how that data is being used are key to understanding the value and, hence, the success of individual data products.

Four Core Principles of Data Mesh



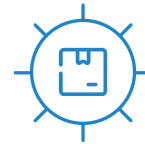
Domain-oriented
Ownership &
architecture



Data as a
product



Self-service data
Infrastructure

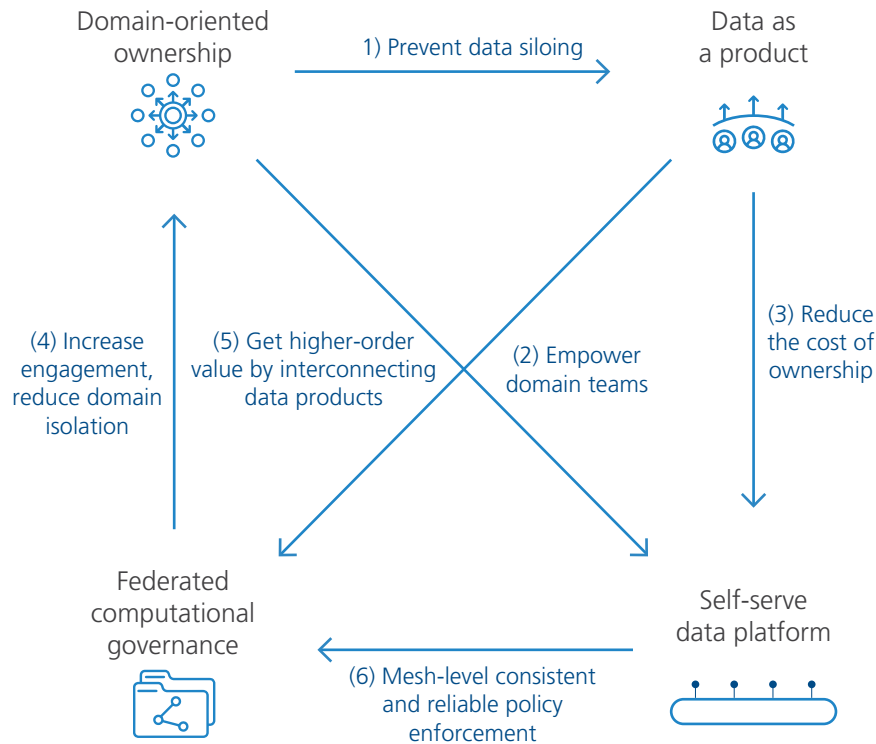


Federated
Computational
Governance

Interplay/Dependency of the Principles

The four principles are collectively necessary and sufficient. They complement each other, and

each addresses new challenges that may arise from others.



For example, decentralized domain-oriented ownership of data can result in data siloing within domains, and this can be addressed by the data as a product principle that demands domains have an organizational responsibility to share their data with product-like qualities inside and outside of their domain. Similarly, the domain ownership of data products can lead to duplicated effort, increased

cost of data product ownership, and lowered data sharing productivity. In this case, the self-serve data platform empowers the cross-functional domain teams in sharing and using data products. The platform objective is to lower the domain teams' cognitive load, reduce unnecessary effort, increase domains' productivity, and lower the total cost of ownership.

Getting Started with Data Mesh

Organizations that are ready to implement data mesh will need help connecting their data sources for a quick win. The following points highlight how:

Connect to data sources where it resides

As you begin your data mesh journey, the first step is to connect to data sources. A key data mesh implementation principle is to connect your data sources by leveraging your existing investments: lakes or warehouses, cloud or on-premise, and structured warehouse or a non-structured lake. Unlike the single-source-of-truth approach to centralize all your data first, you're leveraging and querying the data where it resides

Create logical domains

After generating connectivity across all the various data sets, the next goal is to create an interface for business and analytics teams to find their data. In data mesh terms, we call that a logical domain. It's called logical because data does not move into a repository where data consumers can access it. Rather, a logical place is created where they can log into a dashboard to see the data that's been made available to them.

All the data resides in the domain alongside domain teams that are empowered to work autonomously. In essence, the focus is on the concept of self-service where data consumers can independently do more on their own.

Enable teams to create data products

When the domain team is provided with access to the data they need, the next step is to teach them how to convert data sets into data products. Then, with a data product, create a library or a catalog of data products.

Creating data products is a powerful capability as you enable your data consumers to move quickly from discovery to ideation and insight. This further helps us quickly create and use data products across the organization.

Business Drivers and Benefits of Data Mesh

The implementation of data mesh promotes organizational agility for organizations that want to thrive in an uncertain economic climate. All organizations need to be able to respond to changes in their environment with a low-cost, high-reward approach. Introducing new data sources, needing to comply with changing regulatory requirements, or meeting new analytics requirements are all drivers that will precipitate changes to an organization's data management activities. Current data management approaches are based on complex and heavily integrated Extract, Transform, and

Load (ETL) processes between operational and analytical systems struggling to change in time to support the business needs in the face of these drivers. The purpose of data mesh is to provide a more resilient approach concerning data to efficiently respond to these changes.

Data Mesh is a socio-technical approach that requires changes to the organization across all three dimensions of people, process, and technology. Organizations that adopt it may spend 70% of their efforts on people and processes and 30% on the technology to enable the future data mesh state.

Data Mesh Challenges

The main challenges of a data mesh stem from the complexities inherent to managing multiple data products (and their dependencies) across multiple autonomous domains.

- **Multi-domain data duplication**

Redundancy, which may occur when the data of one domain is repurposed to serve the business needs of another domain, could potentially impact resource utilization and data management costs.

- **Federated data governance and quality assurance**

Different domains may require different data governance tools, which must be taken into account when data products and pipelines are shared commodities. The resulting deltas must be identified and federated.

Here are the key considerations:

- **Change management**

Decentralizing data management to adopt a data mesh approach requires significant change management in highly centralized data management practices.

- **Cost and risk**

Existing data and analytics tools should be adapted and augmented to support a data mesh architecture. Establishing a data management infrastructure to support a data mesh, including data integration, virtualization, preparation, masking, governance, orchestration, cataloging, and delivery, can be a very large, costly, and risky undertaking.

- **Cross-domain analytics**

An enterprise-wide data model must be defined to consolidate the various data products and make them available to authorized users in one central location.

Data Mesh vs. Data Lake

A decorative horizontal bar with a gradient from blue to yellow.

The data lake is a technology approach, whose main objective has traditionally been as a single repository to move data to in as simple a manner as possible, where the central team is responsible for managing it. While data lakes can provide significant business value, they also suffer from several issues. The primary issue is that once data is moved to the lake it loses context, for example, we may have many files containing a definition of the customer, one from a logistics system, one from payments, and one from marketing. Which one should we use? Furthermore, data in the data lake will not have been pre-processed, so data issues will inevitably arise. The data consumer will then

typically have to liaise with the data lake team to understand and resolve data issues, which becomes a significant bottleneck to using the data to answer the initial business question.

In comparison, data mesh is more than just technology. It combines both technology and organizational aspects including the idea of data ownership, data quality, and autonomy. So consumers of data have a clear line of sight around data quality and ownership while issues can be discovered and resolved much more efficiently. Ultimately, this data can be used and trusted.

Data Mesh vs. Data Fabric

Data fabric concentrates on a collection of various technological capabilities that collaborate to produce an interface for the end-users that consume data. Many of the supporters of data fabric espouse automation, through technologies like ML, of many of the data management tasks to simplify end users' access to data. For simple data usage, there is some value in this. However, for more complex situations or where business knowledge needs to be integrated into the data, the limitations of the data fabric will become apparent.

Arguably data fabric could be used as part of a data mesh self-serve platform, where data fabric exposes data to the domains that can then embed their business knowledge into a resulting data product.

As Darnell-Kanal, professor of Computer Science, College Park Daniel Abadi, University of Maryland says, the difference between a data fabric and a data mesh is not obvious. He further advises that "ultimately, an optimal solution will likely take the best ideas from each of these approaches."

Data Mesh Use Cases

Data mesh supports many different operational and analytical use cases, across multiple domains. Here are a few examples:

- **Customer 360-degree view** supports customer care in reducing average handle time, increasing first contact resolution, and improving customer satisfaction. A single view of the customer may also be deployed by marketing to undertake predictive churn modeling or next-best-offer decisions.
- **Data privacy management** protects customer data by complying with ever-emerging regional data privacy laws, like Virginia Consumer Data Protection Act (VCDPA), before making it available to data consumers in the business domains.
- **IoT device monitoring** provides product teams with insights into edge device usage patterns, to continually improve product adoption and profitability.
- **Federated data preparation** enables domains to provide quality and trusted data for their analytics workloads.
- **Hyper-segmentation** enables marketing teams to deliver the right campaign to the right customer, at the right time, and via the right channel.

Data Mesh: What's in it for The Business?

Banking/financial firms: undergoing rapid digital transformation

Without question, banks are investing heavily in digital and mobile technologies, from Artificial Intelligence (AI) powered virtual assistants to 5G apps that make it easy to conduct most, if not all, of your personal and business operations through your mobile phone.

Faced with fierce industry competition, leading banks are developing ways to keep their customers engaged and protected by offering the right products and services at the right time. The ML and AI engines that drive these solutions need data to generate optimal recommendations, but at most

large banks, the data is distributed across multiple disconnected datasets. Moving all of this data into one place isn't going to work, as projects like that take years. By that point, the target customers will have moved on to a new bank or AI-powered finance app.

By shifting to the data mesh approach, banks can feed their AI and ML models the data they need in a secure, efficient manner. And they can make this shift in a matter of months, not years. Before long, they'll be regularly offering customers the products and services they need, when they need them.

Life sciences and pharma: driving discovery

As COVID research has made clear, pharmaceutical companies, university research groups, and government agencies need to collaborate and share data among themselves and with various labs and testing facilities. If all this data is stored in different siloed systems, these organizations aren't going to be able to coordinate their efforts or share

information effectively. But, if companies and research groups adopt the data mesh approach and start to treat datasets as products that can be shared both internally and externally, the potential for discoveries, research findings, and collaborative business partnerships grows tremendously.

SOPHiA GENETICS is a great example. The company has created a global data-sharing network with customers from over 780 institutions in more than 70 countries. They're launching their own business data mesh, which enables them to connect regions and data sources while adhering to compliance laws and regulations. Another example is healthcare technology leader EMIS. The company collects

recently reported COVID-19 symptoms from a wide range of sources and makes them available to researchers in real time. This wouldn't be possible with a centralized, single source of truth model. Data mesh isn't merely changing things for data scientists and infrastructure teams but is opening entirely new opportunities for the business.

Supply chain logistics: supporting the digital twin

Across industries, companies are adopting virtual distribution centers and other digital tools to help them manage complex, distributed networks of assets. Supply chain digital twins simulate and project what is happening and what might happen along your supply chain. They're incredibly valuable, but like the customer-facing AI engines at banks, they only work effectively if they can access relevant, recent data. You need to constantly feed them information to make that possible. If your approach to feeding these models is to move all the datasets from the relevant parties into one

place, you probably won't have a supply chain to manage anymore. The whole operation will take too long. A data mesh that allows these parties to retain control over their datasets, while also making data easily and rapidly accessible to other groups, is a far more efficient way to operate.

These are just a few examples from a handful of industries and application spaces. As data mesh interest and adoption grow, more companies will discover exciting new possibilities.

About the Author

A decorative horizontal bar with a gradient from blue to yellow.

Dharmender Verma



Dharmender has 20+ years of experience in the IT industry and 15 years in enterprise-level data warehousing design using data governance, dimensional modelling, data integration (ETL), and BI analytics. Currently he is positioned as a DW Architect for the IFSCA project. He has previously held various roles like Technical/Solution/BI Architect, Manager, and Team Lead for organizations such as Accenture, Hewlett Packard, Tech Mahindra (Formerly Satyam Computers Services Ltd.), Oracle Financial Services (Formerly I-Flex Solutions).

LTIMindtree is a global technology consulting and digital solutions company that enables enterprises across industries to reimagine business models, accelerate innovation, and maximize growth by harnessing digital technologies. As a digital transformation partner to more than 700 clients, LTIMindtree brings extensive domain and technology expertise to help drive superior competitive differentiation, customer experiences, and business outcomes in a converging world. Powered by 81,000+ talented and entrepreneurial professionals across more than 30 countries, LTIMindtree — a Larsen & Toubro Group company — combines the industry-acclaimed strengths of erstwhile Larsen and Toubro Infotech and Mindtree in solving the most complex business challenges and delivering transformation at scale. For more information, please visit www.ltimindtree.com.