O'REILLY®

Data Mesh

Delivering Data-Driven Value at Scale

Zhamak Dehghani
Foreword by Martin Fowler

O'REILLY®

Data Mesh

We're at an inflection point in data, where our data management solutions no longer match the complexity of organizations, the proliferation of data sources, and the scope of our aspirations to get value from data with AI and analytics. In this practical book, author Zhamak Dehghani introduces data mesh, a decentralized sociotechnical paradigm drawn from modern distributed architecture that provides a new approach to sourcing, sharing, accessing, and managing analytical data at scale.

Dehghani guides practitioners, architects, technical leaders, and decision makers on their journey from traditional big data architecture to a distributed and multidimensional approach to analytical data management. Data mesh treats data as a product, considers domains as a primary concern, applies platform thinking to create self-serve data infrastructure, and introduces a federated computational model of data governance.

- Get a complete introduction to data mesh principles and its constituents
- Design a data mesh architecture
- Guide a data mesh strategy and execution
- Navigate organizational design to a decentralized data ownership model
- Move beyond traditional data warehouses and lakes to a distributed data mesh

"A thorough and crucially needed overview of data as a product, including the cultural, process, technology, and team changes required to get there."

-Manuel Pais
Coauthor of *Team Topologies*

"Zhamak's detailed approach in *Data Mesh* makes this new concept clear and useful."

-Gwen Shapira
Cofounder and CPO at Nile Platform; Author of *Kafka: The Definitive Guide*

Zhamak Dehghani is a director of technology at Thoughtworks, focusing on distributed systems and data architecture in the enterprise. She's a member of multiple technology advisory boards including Thoughtworks. Zhamak is an advocate for decentralization of all things, including architecture, data, and ultimately power. She is the founder of data mesh.

SOFTWARE ARCHITECTURE

US $69.99 CAN $87.99
ISBN: 978-1-492-09239-1

Twitter: @oreillymedia
linkedin.com/company/oreilly-media
youtube.com/oreillymedia

Data Mesh

Delivering Data-Driven Value at Scale

Zhamak Dehghani

Beijing • Boston • Farnham • Sebastopol • Tokyo

O'REILLY®

Data Mesh
by Zhamak Dehghani

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: Melissa Duffield
Development Editor: Gary O'Brien
Production Editor: Beth Kelly
Copyeditor: Charles Roumeliotis
Proofreader: Kim Wimpsett

Indexer: Potomac Indexing, LLC
Interior Designer: David Futato
Cover Designer: Karen Montgomery
Illustrator: Kate Dullea

March 2022: First Edition

Revision History for the First Edition
2022-03-08: First Release
2022-07-15: Second Release

See *http://oreilly.com/catalog/errata.csp?isbn=9781492092391* for release details.

978-1-098-11276-9

[LSI]

CHAPTER 15
Strategy and Execution

The essence of strategy is choosing to perform activities differently than rivals do.
—Michael E. Porter

Is your organization adopting a data-oriented strategy? Is it planning to offer distinct value to its customers and partners using data and insights? Does your organization have a vision to use data to perform a different set of activities or its current activities differently? Is your organization curious about data mesh as a foundation for implementing this strategy?

If you answered yes to these questions, this chapter is for you.

In this chapter I describe the relationship between data mesh and an overall data strategy, and how to approach its execution through an iterative, value-driven, and evolutionary approach.

But first, a quick self-assessment.

Should You Adopt Data Mesh Today?

Before you proceed to the rest of this chapter, you may ask yourself, is data mesh the right choice for my organization? Or more importantly, is data mesh the right choice now?[1]

Figure 15-1 depicts my perspective on these questions. I suggest that organizations that fall in the colored section of the spider graph—scoring medium or high on each of the assessment axes—are best positioned to adopt data mesh successfully at this

1 "Now" is when this book is published and implies an early state of maturity for technologies that natively support data mesh.

point in time. You can use the following criteria to self-assess whether data mesh is the right choice for your organization now.

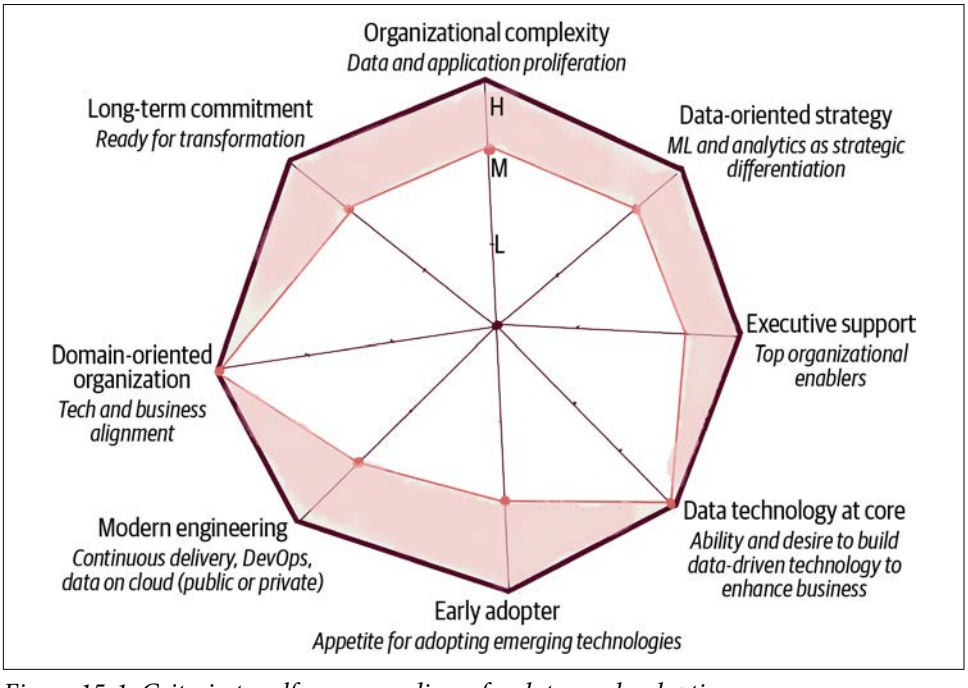

Figure 15-1. Criteria to self-assess readiness for data mesh adoption

Organizational complexity

Data mesh is a solution for organizations that experience scale and complexity, where existing data warehouse or lake solutions have become blockers in their ability to get value from data at scale and across many functions of their business, in a timely fashion and with less friction.

If your organization has not reached such a level of complexity and the existing data and analytics architectures meet your needs, data mesh will not offer you much value.

Organizations with a large proliferation of data sources and use cases that score moderate or high on data complexity benefit from data mesh. For example, enterprises or scaleups—fast-growing startups—whose business encapsulates many different data domains and a large proliferation of data types collected from many sources and that have a business strategic differentiator heavily dependent on ML and analytical use cases are great candidates for data mesh. Enterprises or scaleups in industries such as retail, banking and finance, insurance, and large tech services fall into this category.

For example, consider a healthcare provider or payer. There is an inherent complexity in the diversity of data. It ranges from clinical visits to patient profiles, lab results, pharmaceuticals, and clinical claims. Each of these categories of data can be sourced from multiple parties in an ecosystem of business domains or independent organizations, hospitals, pharmacies, clinics, labs, etc. The data is being used for many use cases: optimized and personalized care models, intelligent insurance authorization, and value-based care.

The primary criteria for adopting data mesh are the intrinsic business complexity and proliferation of data sources and data use cases.

Data-oriented strategy

Data mesh is a solution for an organization planning to get value from data at scale. It requires the commitment of product teams and business units to imagine using intelligent decision making and actions in their applications and services. Such a commitment is possible only when the organization identifies data enabling ML and analytics as a strategic differentiator.

A data-oriented strategy then translates down to data-driven applications and services, which then turns into teams' commitment to data.

Without a strategic vision around data usage, it would be difficult to motivate domain teams and business domains to take on the extra responsibility of data sharing.

Executive support

Data mesh introduces change, and with that comes resistance to change. Any change of such scale experiences detractors and blockers. It will be necessary to make tough decisions between the progress of the platform versus building point solutions to meet a deadline. It demands motivating the organization to change how people work.

All this calls for executive support and top-down engagement of leaders. In my experience, the most successful data mesh implementations have had the backing of C-level executives.

Data technology at core

Implementing data mesh builds on a new relationship between technology and business. It builds on the pervasive application of data in each business domain. Organizations that have data technology at their core use data and AI as a competitive advantage, as the means to expand and reshape their business and not merely as an enabler.

Such organizations have the desire and ability to build and create the technology needed to embed data sharing and consumption at the core of each business function.

In contrast, organizations that see technology as a supporter of the business and not at the core often delegate and externalize technology capabilities to external vendors and expect to buy and plug in a ready-made solution for their business needs. Such organizations are not yet positioned to adopt data mesh.

Early adopter

At the time of writing this book, data mesh is enjoying its early years of adoption by innovators and lead adopters. Early adopters of new innovations must have certain characteristics: they are venturesome opinion leaders[2] in their industry. They have an appetite for adopting emerging technologies, such as data mesh.

Early adopters of a novel approach, particularly one as multidimensional and pervasive as data mesh, demand a spirit of experimentation, taking risks, failing fast, learning, and evolving. Companies that adopt data mesh at this point in time must be willing to begin from the first principles, adapt them to their context, and learn and evolve.

On the contrary, organizations that don't take risks and like to only adopt well-tested, refined, and prescribed playbooks and fall into the late adopter category may need to just wait for some time.

Modern engineering

Data mesh builds on a foundation grounded in modern software engineering practices. Continuous and automated delivery of software, DevOps practices, distributed architecture, computational policies, and availability of modern data storage and processing stacks—on private or public clouds. Without solid engineering practices and access to open and modern data tooling, it will be challenging to bootstrap data mesh from the ground up. Data mesh shares prerequisites for microservices[3] and requires modern data and analytics technology and practices.

Additionally, data mesh works with modern technology stacks that are API driven and easy to integrate and aim to enable many smaller teams, as opposed to one large, centralized team. Technologies that enforce centralized data modeling, centralized control of data, and centralized storage of data don't lend themselves to a successful data mesh implementation. They become a bottleneck and conflict with the distributed nature of data mesh.

Domain-oriented organization

Data mesh assumes that the adopters either are modern digital businesses with a level of technology and business alignment or are on the journey toward it. It

2 Everett M. Rogers, *Diffusion of Innovations*, 1.

3 Martin Fowler, "Microservice Prerequisites", (2014).

assumes that organizations are designed based on their business domains, where each business domain has dedicated technical teams who build and support the digital assets and products that serve the business domain. It assumes that technology and business teams are in close collaboration to use technology to enable, reshape, and expand the business.

For example, a modern bank has dedicated technology groups for digital core banking, digital processing of loans, digital commercial and personal banking, digital credit and risk, etc. There is a close alignment between technology, technical teams, and the business domains.

In contrast, organizations with a centralized IT unit that shares people (resources) based on projects across all business units, without continuous and long-term ownership of technical assets for each business domain, are not a good fit for data mesh.

Domain-oriented data sharing scales only if there is a domain-oriented tech and business alignment in place.

Long-term commitment
: As you will see in upcoming chapters, the adoption of data mesh is a transformation and a journey. While it can be started with a limited scope and an organizational footprint of a small number of domains, the benefits are compounded when exploited at scale. Hence, you need to be able to create an organizational commitment to the vision and embark on a transformation journey.

  Data mesh cannot be implemented as one or two one-off projects.

If you see yourself aligned with the previous criteria, let's go ahead and look at how to execute a data strategy using data mesh.

Data Mesh as an Element of Data Strategy

To understand how data mesh fits into a larger data strategy, let's look at our hypothetical company, Daff, Inc.

Daff's data strategy is to create an *intelligent platform* that connects artists and listeners through an immersive artistic experience. This experience is *continuously* and *rapidly improved* using data and ML embedded in each and every interaction of the listeners and artists on the platform. This implies capturing data from every single touchpoint on the platform and augmenting every platform feature using ML derived from the engagement data. For example, the platform's better playlist predictions lead to more listeners with higher engagement, which then leads to more data and hence better predictions—creating a positive self-reinforcing loop for the business.