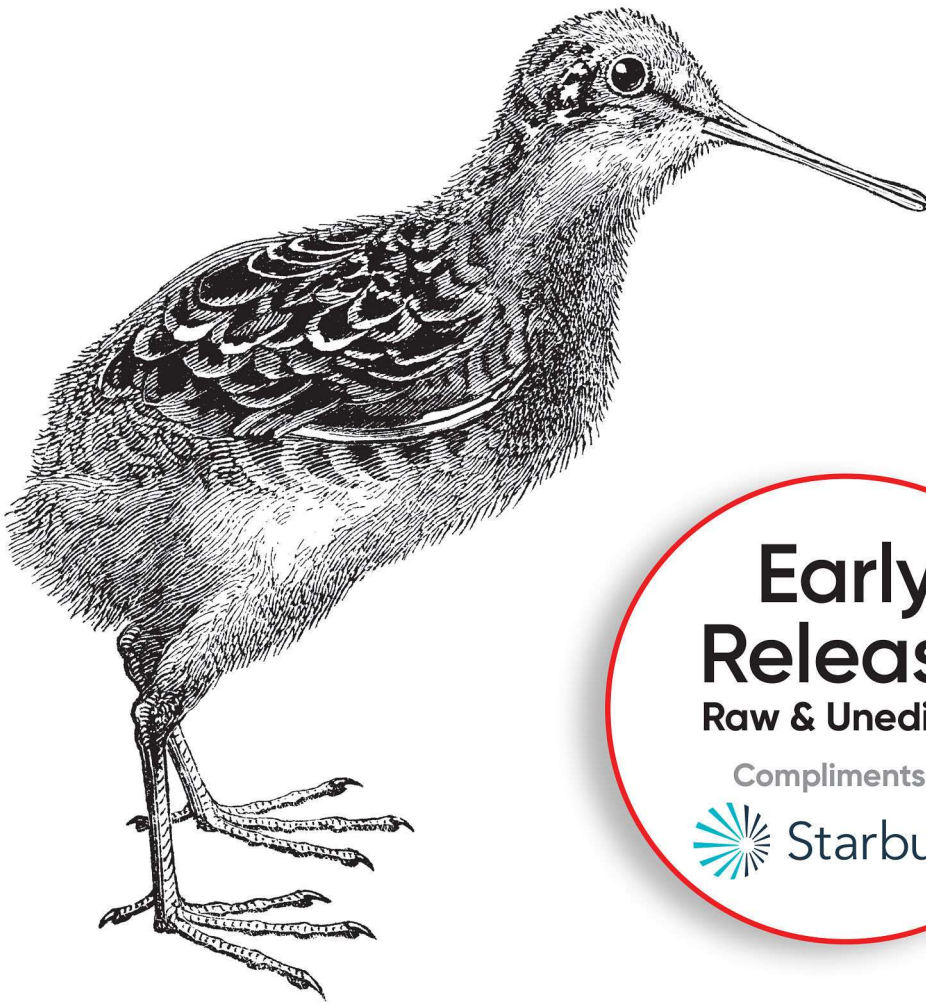


**O'REILLY®**

# Data Mesh

Delivering Data-Driven Value at Scale



**Early  
Release**  
Raw & Unedited

Compliments of



Starburst

Zhamak Dehghani

# Data Mesh Resource Center

Compliments of



Class is in session: [starburst.io/datamesh](https://starburst.io/datamesh)

Join your peers and enjoy exclusive access  
to educational Data Mesh content.

On-demand talks, panel discussions  
featuring Zhamak Dehghani, and more!

---

# Data Mesh

## *Delivering Data-Driven Value at Scale*

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.

*Zhamak Dehghani*

Beijing • Boston • Farnham • Sebastopol • Tokyo

**O'REILLY®**

## Data Mesh

by Zhamak Dehghani

Copyright © 2022 Zhamak Dehghani. All rights reserved.

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](mailto:corporate@oreilly.com).

**Acquisitions Editor:** Melissa Duffield

**Development Editor:** Gary O'Brien

**Production Editor:** Beth Kelly

**Interior Designer:** David Futato

**Cover Designer:** Karen Montgomery

**Illustrator:** Kate Dullea

January 2022: First Edition

### Revision History for the Early Release

2021-06-18: First Release

2021-07-28: Second Release

2021-09-07: Third Release

2021-10-07: Fourth Release

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

The O'Reilly logo is a registered trademark of O'Reilly Media, Inc. *Data Mesh*, the cover image, and related trade dress are trademarks of O'Reilly Media, Inc.

The views expressed in this work are those of the authors, and do not represent the publisher's views. While the publisher and the authors have used good faith efforts to ensure that the information and instructions contained in this work are accurate, the publisher and the authors disclaim all responsibility for errors or omissions, including without limitation responsibility for damages resulting from the use of or reliance on this work. Use of the information and instructions contained in this work is at your own risk. If any code samples or other technology this work contains or describes is subject to open source licenses or the intellectual property rights of others, it is your responsibility to ensure that your use thereof complies with such licenses and/or rights.

This work is part of a collaboration between O'Reilly and Starburst Data. See our [statement of editorial independence](#).

978-1-492-09232-2

[FILL IN]

---

# Table of Contents

---

## Part I. Why Data Mesh?

<b>1. The Inflection Point. ....</b>	<b>13</b>
Great Expectations of Data	15
The Great Divide of Data	16
Operational Data	17
Analytical Data	18
Analytical and Operational Data Misintegration	19
Scale, Encounter of a New Kind	20
Beyond Order	21
Approaching the Plateau of Return	22
Recap	22
<b>2. After The Inflection Point. ....</b>	<b>25</b>
Embrace Change in a Complex, Volatile and Uncertain Business	
Environment	27
Align Business, Tech and Now Analytical Data	27
Close The Gap Between Analytical and Operational Data	28
Localize Data Change to Business Domains	30
Reduce Accidental Complexity of Pipelines and Copying Data	31
Sustain Agility in the Face of Growth	31
Remove Centralized and Monolithic Bottlenecks of the Lake or the	
Warehouse	32
Reduce Coordination of Data Pipelines	33
Reduce Coordination of Data Governance	33
Enable Autonomy	34
Increase the Ratio of Value from Data to Investment	35

Abstract Technical Complexity with a Data Platform	35
Embed Product Thinking Everywhere	35
Go Beyond The Boundaries	36
Recap	36
<b>3. Before The Inflection Point. ....</b>	<b>39</b>
Evolution of Analytical Data Architectures	40
First Generation: Data Warehouse Architecture	40
Second Generation: Data Lake Architecture	41
Third Generation: Multimodal Cloud Architecture	43
Characteristics of Analytical Data Architecture	44
Monolithic	46
Monolithic Architecture	46
Monolithic Technology	48
Monolithic Organization	48
The complicated monolith	50
Technically-Partitioned Architecture	53
Activity-oriented Team Decomposition	54
Recap	55

---

## Part II. What is Data Mesh

<b>4. Principle of Domain ownership. ....</b>	<b>65</b>
Apply DDD's Strategic Design to Data	66
Domain Data Archetypes	68
Source-aligned Domain Data	69
Aggregate Domain Data	71
Consumer-aligned Domain Data	71
Transition to Domain Ownership	72
Push Data Ownership Upstream	72
Define Multiple Connected Models	72
Embrace the Most Relevant Domain, and Don't Expect the Single Source of Truth	73
Hide the Data Pipelines as Domains' Internal Implementation	74
Recap	74
<b>5. Principle of Data as a Product. ....</b>	<b>75</b>
Apply Product Thinking to Data	77
Baseline usability characteristics of a data product	79
Transition to Data as a Product	86
Include Data Product Ownership in Domains	87

Recap	92
<b>6. Principle of Self-Serve Data Platform. ....</b>	<b>93</b>
Data Mesh Platform, Compare and Contrast	95
Serving Autonomous Domain-oriented Teams	96
Managing Autonomous and Interoperable Data Products	97
A Continuous Platform of Operational and Analytical Capabilities	97
Designed for Generalists Majority	97
Favoring Decentralized Technologies	98
Domain Agnostic	98
Data Mesh Platform Thinking	99
Enable Autonomous Teams to Get Value from Data	101
Exchange Value with Autonomous and Interoperable Data Products	103
Accelerate Exchange of Value by Lowering the Cognitive Load	104
Scale out Data Sharing	105
Support a Culture Of Embedded Innovation	106
Transition To Self-serve Data Mesh Platform	107
Design the APIs and Protocols First	107
Prepare for Generalists Adoption	107
Create Higher Level APIs to Manage Data Products	108
Converge Data and Operational Platforms, Where Possible	108
Build Experiences, not Mechanisms	109
Begin with the Simplest Foundation, then Harvest to Evolve	109
Recap	110
<b>Prospective Table of Contents (Subject to Change). ....</b>	<b>111</b>

---

# Why Data Mesh?

*By doubting we are led to question, by questioning we arrive at the truth.*

—Peter Abelard

Data Mesh is a new approach in sourcing, managing, and accessing data for analytical use cases at *scale*. Let's call this class of data, analytical data. Analytical data is used for predictive or diagnostic use cases. It is the foundation for visualizations and reports that provide insights into the business. It is used to train machine learning models that augment the business with data-driven intelligence. It is the essential ingredient for organizations to move from intuition and gut-driven decision-making to taking actions based on observations and data-driven predictions. Analytical data is what powers the software and technology of the future. It enables a technology shift from human-designed rule-based algorithms to data-driven machine-learned models. Analytical data is becoming an increasingly critical component of the technology landscape.



The phrase data in this writeup, if not qualified, refers to analytical data. Analytical data serves reporting and machine learning training use cases.

Data Mesh calls for a fundamental shift in our assumptions, architecture, technical solutions, and social structure of our organizations, in how we manage, use, and own analytical data.