

O'REILLY®

Deutsche
Ausgabe

Data Mesh

Eine dezentrale Datenarchitektur entwerfen



Zhamak Dehghani

Vorwort von Martin Fowler

Übersetzung von Jochen Christ und Simon Harrer

Copyright und Urheberrechte:

Die durch die dpunkt.verlag GmbH vertriebenen digitalen Inhalte sind urheberrechtlich geschützt. Der Nutzer verpflichtet sich, die Urheberrechte anzuerkennen und einzuhalten. Es werden keine Urheber-, Nutzungs- und sonstigen Schutzrechte an den Inhalten auf den Nutzer übertragen. Der Nutzer ist nur berechtigt, den abgerufenen Inhalt zu eigenen Zwecken zu nutzen. Er ist nicht berechtigt, den Inhalt im Internet, in Intranets, in Extranets oder sonst wie Dritten zur Verwertung zur Verfügung zu stellen. Eine öffentliche Wiedergabe oder sonstige Weiterveröffentlichung und eine gewerbliche Vervielfältigung der Inhalte wird ausdrücklich ausgeschlossen. Der Nutzer darf Urheberrechtsvermerke, Markenzeichen und andere Rechtsvorbehalte im abgerufenen Inhalt nicht entfernen.

Data Mesh

Eine dezentrale Datenarchitektur entwerfen

Zhamak Dehghani

*Deutsche Übersetzung von
Jochen Christ & Simon Harrer*

O'REILLY®

Zhamak Dehghani

Lektorat: Alexandra Follenius

Übersetzung: Jochen Christ, Simon Harrer

Copy-Editing: Sibylle Feldmann, www.richtiger-text.de

Satz: Jörg Liedtke, www.liedtke.me

Herstellung: Stefanie Weidner

Umschlaggestaltung: Karen Montgomery, Michael Oréal, www.oreal.de

Druck und Bindung: mediaprint solutions GmbH, 33100 Paderborn

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

ISBN:

Print 978-3-96009-207-0

PDF 978-3-96010-724-8

ePub 978-3-96010-725-5

mobi 978-3-96010-726-2

1. Auflage 2023

Translation Copyright für die deutschsprachige Ausgabe © 2023 dpunkt.verlag GmbH

Wieblinger Weg 17

69123 Heidelberg

Authorized German translation of the English edition of *Data Mesh*, ISBN 9781492092391

© 2022 Zhamak Dehghani. This translation is published and sold by permission of O'Reilly Media, Inc., which owns or controls all rights to publish and sell the same.

Dieses Buch erscheint in Kooperation mit O'Reilly Media, Inc. unter dem Imprint »O'REILLY«. O'REILLY ist ein Markenzeichen und eine eingetragene Marke von O'Reilly Media, Inc. und wird mit Einwilligung des Eigentümers verwendet.

Hinweis:

Dieses Buch wurde auf PEFC-zertifiziertem Papier aus nachhaltiger Waldwirtschaft gedruckt. Der Umwelt zuliebe verzichten wir zusätzlich auf die Einschweißfolie.



Schreiben Sie uns:

Falls Sie Anregungen, Wünsche und Kommentare haben, lassen Sie es uns wissen: komentar@oreilly.de.

Die vorliegende Publikation ist urheberrechtlich geschützt. Alle Rechte vorbehalten. Die Verwendung der Texte und Abbildungen, auch auszugsweise, ist ohne die schriftliche Zustimmung des Verlags urheberrechtswidrig und daher strafbar. Dies gilt insbesondere für die Vervielfältigung, Übersetzung oder die Verwendung in elektronischen Systemen.

Es wird darauf hingewiesen, dass die im Buch verwendeten Soft- und Hardware-Bezeichnungen sowie Markennamen und Produktbezeichnungen der jeweiligen Firmen im Allgemeinen warenzeichen-, marken- oder patentrechtlichem Schutz unterliegen.

Alle Angaben und Programme in diesem Buch wurden mit größter Sorgfalt kontrolliert. Weder Autorin noch Verlag noch Übersetzer können jedoch für Schäden haftbar gemacht werden, die in Zusammenhang mit der Verwendung dieses Buches stehen.

5 4 3 2 1 0

Für Dad

Dein Licht leuchtet weiter.

Vorwort	15
Vorbemerkung der Übersetzer	17
Einleitung	19
Prolog: Fallstudie	27

Teil I Was ist Data Mesh?

1 Data Mesh im Überblick	45
Ergebnisse	45
Veränderungen	46
Prinzipien	48
Domain Ownership	48
Data as a Product	49
Self-Serve Data Platform	50
Federated Computational Governance	50
Zusammenspiel der Prinzipien	51
Data Mesh kurz erklärt	52
Daten	53
Operative Daten	53
Analytische Daten	54
Entstehungsgeschichte	55
2 Das Prinzip Domain Ownership	57
Domain-driven Design	59
Strategisches Design	60
Archetypen von Domänendaten	63
Quellenorientierte Domänendaten	64
Aggregierte Domänendaten	65
Konsumentenorientierte Domänendaten	66

Einführung von Domain Ownership	67
Verantwortung wandert flussaufwärts.	67
Verknüpfung von Modellen.	67
Mythos Single Source of Truth	68
Daten-Pipelines als Implementierungsdetail	69
Zusammenfassung	70
3 Das Prinzip Data as a Product	71
Product Thinking	73
Grundlegende Usability-Attribute von Datenprodukten.	75
Einführung von Data as a Product	84
Domänen verantworten Datenprodukte	84
Eine neue Sprache	85
Daten als Produkt statt als Asset	86
Vertraue, aber prüfe nach	87
Daten und Logik als Einheit.	87
Zusammenfassung	88
4 Das Prinzip Self-Serve Data Platform	91
Data-Mesh-Plattform im Vergleich.	93
Autonome domänenorientierte Teams als Zielgruppe.	93
Autonome und interoperable Datenprodukte als Kernkonzept.	95
Operative und analytische Funktionen in einer integrierten Plattform.	95
Generalisten als Hauptnutzende	96
Dezentrale Technologien	97
Domänenagnostisch.	98
Plattform Thinking.	98
Wertschöpfung durch autonome Teams	101
Wertschöpfung mit autonomen und interoperablen Datenprodukten.	102
Reduktion des Cognitive Load.	103
Skalierung der Bereitstellung von Daten	105
Förderung einer Innovationskultur	106
Einführung einer Self-Serve Data Platform	107
API first	107
Optimierung für Generalisten	107
Simplify	108
High-Level-APIs für Datenprodukte	108
Nutzerorientierung statt Feature-Orientierung	109
Klein anfangen und mitwachsen	109
Zusammenfassung	110

5	Das Prinzip Federated Computational Governance	111
	Systems Thinking	114
	Gleichgewicht zwischen Autonomie und Interoperabilität	115
	Dynamische Topologie als Standardzustand	119
	Automatisierung und verteilte Architektur	119
	Föderalismus	120
	Föderales Team	121
	Leitlinien	123
	Policies	126
	Anreize	127
	Automatisierung	128
	Standards als Code	129
	Policies als Code	130
	Automatisierte Tests	131
	Automatisiertes Monitoring	131
	Einführung von Federated Computational Governance	131
	Delegation der Verantwortlichkeit an Domänen	132
	Integration der Policy-Ausführung in jedes Datenprodukt	132
	Automatisierung und Monitoring der Policies	133
	Bestimmung von Polysemen	133
	Messung des Netzwerkeffekts	133
	Mut zu Veränderung	134
	Zusammenfassung	134

Teil II Warum Data Mesh?

6	Der Wendepunkt	139
	Die großen Erwartungen	140
	Die große Kluft	142
	Skalierung: Begegnung einer neuen Art	144
	Nichts ist so beständig wie der Wandel	145
	Diskrepanz zwischen Investition und Rendite	146
	Zusammenfassung	147
7	Nach dem Wendepunkt	149
	Auf Veränderungen angemessen reagieren	150
	Alignment von Fachlichkeit, Technologie und analytischen Daten	151
	Integration der operativen und analytischen Welt	152
	Dezentralisierung von Datenänderungen	154
	Beseitigung unbeabsichtigter Komplexität	154

Agilität trotz Wachstum	155
Beseitigung zentralisierter und monolithischer Komponenten	155
Reduzierung des Koordinierungsaufwands von Pipelines	156
Reduzierung des Koordinationsaufwands für Data Governance	157
Förderung der Autonomie	158
Verbesserung des Kosten-Nutzen-Verhältnisses	159
Komplexitätsreduktion durch die Datenplattform	160
Product Thinking überall	160
Organisationsübergreifender Datenzugriff	161
Zusammenfassung	161
8 Vor dem Wendepunkt	165
Evolution der analytischen Datenarchitekturen	165
Erste Generation: Data-Warehouse-Architektur	166
Zweite Generation: Data-Lake-Architektur	167
Dritte Generation: multimodale Cloud-Architektur	169
Eigenschaften der analytischen Datenarchitektur	170
Monolithisch	171
Zentralisiert	175
Technologieorientiert	177
Zusammenfassung	181

Teil III Wie entwirft man eine Data-Mesh-Architektur?

9 Logische Architektur	185
Schnittstellen für die Bereitstellung analytischer Domänenendaten.	188
Design der operativen Schnittstelle	190
Design der analytischen Datenschnittstelle	190
Domänenübergreifende Abhängigkeiten von Analysedaten	191
Datenprodukt als Architekturquantum	192
Strukturelle Komponenten eines Datenprodukts	193
Bereitstellung und Konsum von Daten	199
Discovery- und Observability-APIs	201
Die drei Ebenen der Datenplattform	201
Dateninfrastruktur-Ebene	203
Datenprodukt-Ebene	204
Mesh-Ebene	204
Beispiel	204
Eingebettete automatisierte Policies	205
Datenprodukt-Sidecar	207

Datenprodukt-Container	208
Control-Port	208
Zusammenfassung	209
10 Datenplattform	213
Entwurf der Plattform anhand von User Journeys	216
User Journey der Data Product Developer	217
Incept, Explore, Bootstrap, Source	219
Build, Test, Deploy, Run	222
Maintain, Evolve, Retire	225
User Journey der Data Product Consumer	227
Incept, Explore, Bootstrap, Source	230
Build, Test, Deploy, Run	231
Maintain, Evolve, Retire	232
Zusammenfassung	232

Teil IV Wie entwirft man Datenprodukte?

11 Entwurf eines Datenprodukts anhand von Affordances	235
Datenprodukt-Affordances	236
Merkmale einer Datenproduktarchitektur	239
Der Einfluss von komplexen adaptiven Systemen auf den Entwurf	240
Emergentes Verhalten aus einfachen lokalen Regeln	241
Kein zentraler Koordinator	242
Zusammenfassung	242
12 Daten konsumieren, transformieren und bereitstellen	243
Daten bereitstellen	243
Anforderungen der Datennutzer	243
Entwurfseigenschaften für das Bereitstellen von Daten	246
Entwurf	259
Daten konsumieren	260
Archetypen von Datenquellen	261
Lokalität des Datenkonsums	265
Entwurf	267
Daten transformieren	268
Programmatische vs. nichtprogrammatische Transformation	269
Datenflussbasierte Transformation	270
Machine Learning als Transformation	271
Zeitabhängige Transformation	272
Entwurf	272
Zusammenfassung	274

13 Daten finden, verstehen und kombinieren	275
Daten finden und verstehen	275
Gefunden werden dank Selbstregistrierung	278
Finden des globalen URI	278
Semantische und syntaktische Modelle verstehen	279
Vertrauen schaffen mit Datengarantien	280
Untersuchung der Form von Daten	284
Lernen mit Dokumentation	284
Entwurf	285
Daten kombinieren	286
Anforderungen	288
Traditionelle Ansätze	289
Entwurf	293
Zusammenfassung	295
14 Daten verwalten, regeln und beobachten	297
Lebenszyklus verwalten	297
Entwurf	298
Manifest	299
Daten regeln	300
Entwurf	301
Standardisierte Policies	302
Integration von Daten und Policies	304
Verknüpfung von Policies	304
Daten beobachten	305
Entwurf	306
Zusammenfassung	309

Teil V Wie führt man Data Mesh ein?

15 Strategie und Umsetzung	313
Sollten Sie Data Mesh schon jetzt einführen?	313
Data Mesh als ein Teil der Datenstrategie	317
Data Mesh Execution Framework	321
Fachlich getriebene Umsetzung	322
Iterative Ende-zu-Ende-Umsetzung	327
Evolutionäre Umsetzung	328
Zusammenfassung	344

16 Organisation und Unternehmenskultur	345
Veränderungen	347
Kultur.	349
Werte	350
Anreize	352
Intrinsische Motivation.	353
Extrinsische Motivation	354
Struktur	355
Annahmen über die Organisationsstruktur	355
Zuschnitt von Datenprodukten	363
Menschen	367
Rollen	367
Weiterbildung.	370
Prozess	372
Wesentliche Prozessänderungen.	374
Zusammenfassung.	375
Index	377

Ich beschäftige mich schon seit mehreren Jahrzehnten mit der Entwicklung von Software für große Unternehmen, und die Verwaltung von Daten war schon immer ein wichtiger Aspekt der Architektur. In den Anfängen meiner Karriere gab es viel Enthusiasmus für ein unternehmensweit einheitliches Datenmodell, das oft in der unternehmensweiten Datenbank gespeichert wurde. Wir haben jedoch schnell gelernt, dass dies zu einer katastrophalen Kopplung führt, wenn eine Vielzahl von Anwendungen auf einen gemeinsamen Datenspeicher zugreift. Aber auch ohne dies gab es größere Probleme. Kernideen eines Unternehmens, wie z.B. »Kunde«, erforderten unterschiedliche Datenmodelle in verschiedenen Domänen. Unternehmensübernahmen machten die Sache noch schwieriger.

Deshalb haben kluge Unternehmen ihre Daten dezentralisiert und Datenspeicherung, -modelle und -verwaltung in die fachlichen Domänen verlagert. Auf diese Weise sind Personen, die die Daten in ihrer Domäne am besten verstehen, auch für die Verwaltung dieser Daten verantwortlich. Sie arbeiten mit anderen Domänen über definierte APIs zusammen. Da diese APIs Logik enthalten können, haben wir mehr Flexibilität bei der Bereitstellung dieser Daten und, was noch wichtiger ist, bei der Weiterentwicklung der Datenverwaltung im Laufe der Zeit.

Während sich dies für das operative Tagesgeschäft immer mehr durchgesetzt hat, blieb die Datenanalyse eine eher zentralisierte Aufgabe. Data Warehouses zielten darauf ab, ein unternehmensweites Repository für kuratierte kritische Informationen bereitzustellen. Doch eine solche zentralisierte Organisationseinheit hatte mit dieser Aufgabe und ihren konkurrierenden Datennutzern zu kämpfen, zumal sie weder die Daten noch die Bedürfnisse ihrer Datennutzer richtig verstanden. Ein Data Lake half, indem er Analysten durch den Zugang zu Rohdaten ermöglichte, näher an die ursprüngliche Quelle heranzukommen. Aber allzu leicht wurde er zu einem Datensumpf mit geringer Datenqualität, mangelhaftem Datenverständnis und unklarer Datenherkunft.

Data Mesh versucht nun, die Erfahrungen, die wir mit operativen Daten gemacht haben, auch auf die Welt der analytischen Daten anzuwenden. Die fachlichen Domänen sind für die Bereitstellung analytischer Daten über APIs verantwortlich,

so wie sie es auch für operative Daten tun. Indem sie ihre Daten als echte Produkte betrachten, kommunizieren sie die Bedeutung und Herkunft der Daten und arbeiten eng mit ihren Nutzern zusammen. Um die damit verbundene Aufgabe bewältigen zu können, muss das Unternehmen eine Plattform für die Erstellung und Bereitstellung dieser Datenprodukte zur Verfügung stellen, zusammen mit einer föderalen Governance-Struktur, um die Kohärenz des Ganzen zu gewährleisten. Bei all dem spielt technische Exzellenz eine wichtige Rolle, damit sich die Plattformen und Produkte schnell weiterentwickeln können, wenn sich die Geschäftsanforderungen ändern.

Data Mesh ist also im Grunde eine recht einfache, vielleicht sogar offensichtliche Anwendung eines gut etablierten Grundsatzes der Datenverwaltung auf die Welt der analytischen Daten. In der Praxis ist es jedoch sehr schwierig, dies zu realisieren, vor allem weil sich so viele Investitionen der Anbieter auf zentralisierte Modelle konzentriert haben. Dies wird noch dadurch erschwert, dass die Praktiken (wie Testen, Abstraktionen und Refactoring) nicht unterstützt werden, von denen die Entwicklerinnen und Entwickler von operativen Systemen wissen, dass sie für gesunde Software unerlässlich sind.

Zhamak war an vorderster Front dabei, beriet unsere Kunden auf dem Weg in die Zukunft, lernte aus ihren Rückschlägen und Erfolgen und ermunterte die Anbieter, die Tools zu entwickeln, die den Aufbau dieser Plattformen erleichtern. Dieses Buch bündelt ihr Wissen und das ihrer Kollegen in dieser frühen, aber wichtigen Phase der weltweiten Einführung von Data Mesh. Ich habe beim Review dieses Buchs viel über die damit verbundenen Herausforderungen gelernt. Ich bin überzeugt, dass alle, die ihre Daten optimal nutzen möchten, in diesem Buch den besten Weg finden, den wir kennen.

– Martin Fowler
Chief Scientist, Thoughtworks

Vorbemerkung der Übersetzer

Wir sind Softwareentwickler. In unseren Projekten arbeiten wir typischerweise in kleinen, autonomen Entwicklungsteams, die für eine bestimmte Domäne zuständig sind. Technisch entwickeln und betreiben wir Anwendungen mit UIs, APIs und operativen Datenbanken in Form von Self-contained Systems (<https://scs-architecture.org/>) in der Cloud. Datenanalysen spielten in unserer täglichen Arbeit lange keine Rolle.

2018 stellte Zhamak Dehghani ihre Idee von Data Mesh vor. Für uns waren ihre beiden Artikel dazu wie eine Art Augenöffner. Data Mesh führt die Ideen von Domain-driven Design, Self-contained Systems, Team Topologies und Product Thinking fort und wendet diese auf die Welt der Daten an.

Wir waren damals Teil eines autonomen Produktteams und haben dann begonnen, die Ideen von Data Mesh bei uns anzuwenden. Wir haben auf Google Big-Query, bereitgestellt vom Datenplattformteam, auf Basis von Domänen-Events Datenprodukte erstellt und darauf Datenanalysen durchgeführt. Dadurch konnten wir als Entwicklungsteam zusammen mit unserem Product Owner eigenständig Hypothesen vor der Umsetzung validieren und den Nutzen nach der Umsetzung messen. Es war unglaublich motivierend, nachweislich sinnvolle Features umsetzen und konkrete Erfolge gemeinsam feiern zu können. Auch wenn wir nur ein einzelnes Entwicklungsteam waren, konnten wir das große Potenzial von Data Mesh für die gesamte Organisation spüren.

Data Mesh ist der nächste logische Schritt in der Dezentralisierung der Softwarearchitektur. Wir wollen diese Idee unterstützen. Deswegen die Übersetzung dieses wegweisenden Buchs, damit sich die Idee auch im deutschsprachigen Raum verbreitet.

– Jochen Christ und Simon Harrer
Zoom, 2022

PS: In eigener Sache möchten wir noch auf unsere Webseite <https://www.datamesh-architecture.com/> verweisen, auf der wir näher auf die Perspektive aus Richtung Anwendungsentwicklung und Softwarearchitektur eingehen.

Einleitung

Data Mesh ist der Impuls, der uns in der Art, wie wir an Daten herangehen, auf einen neuen Kurs bringt: wie wir uns Daten vorstellen, wie wir sie erfassen und weitergeben und wie wir aus ihnen Nutzen generieren – im großen Maßstab und im Bereich der Datenanalyse und der künstlichen Intelligenz. Dieser neue Kurs führt uns weg von der Zentralisierung von Daten und deren Ownership hin zu einem dezentralen Modell. Dieser neue Kurs trägt der Komplexität unserer Organisationen, ihrem schnellen Wandel und ihrem kontinuierlichen Wachstum Rechnung. Er zielt darauf ab, selbst große Organisationen in die Lage zu versetzen, trotz des Durcheinanders und der organisatorischen Komplexität einen Mehrwert aus Daten zu ziehen.

Wenn wir auf die Geschichte unserer Branche zurückblicken, haben wir schon einmal einen solchen Impuls erhalten. Die Entstehung von Unix und seiner Philosophie »Schreibe Programme so, dass sie nur eine Aufgabe erledigen und diese gut machen. Schreibe Programme so, dass sie zusammenarbeiten ...« war vielleicht der Schmetterling, der mit seinen Flügeln schlug und die Voraussetzungen dafür schuf, dass wir Jahrzehnte später die Komplexität im Herzen von *Software* durch verteilte Architektur, serviceorientiertes Design, Kommunikation über Standard-APIs und autonome Domänenteams bewältigen konnten. Ich hoffe, dass Data Mesh die Voraussetzung für einen neuen Weg zur Bewältigung der Komplexität im Herzen von *Daten* in dem Bereich schafft, der sie am meisten benötigt, nämlich Datenanalyse und künstliche Intelligenz.

Ich habe die These von Data Mesh im Jahr 2018 formuliert, nachdem ich in großen und technologisch fortschrittlichen Unternehmen, die erhebliche Investitionen in ihre Datentechnologien getätigt hatten, häufig auftretende Fehler bei der Wertschöpfung aus Daten beobachtet hatte. Die beobachteten Schwierigkeiten bei der Skalierung von Systemen und der Organisation des Datenmanagements, um ihre ehrgeizigen Datenziele zu erreichen, führten dazu, dass ich die jahrzehntelangen Annahmen über die Art und Weise, wie wir aus Daten Wert schöpfen, infrage stellte: Wir sammeln sie, wir speichern sie zentral, wir beauftragen ein Datenteam mit ihrer Verwaltung, und dann lassen wir sie auf eine Vielzahl von Anwendungsfällen los. Diese Annahmen mussten überarbeitet werden.

Die Ideen hinter Data Mesh habe ich etwa zur gleichen Zeit auf einer O'Reilly-Konferenz in New York vorgestellt. Ich nannte den Vortrag »Beyond the Lake« (<https://oreil.ly/O3hbf>), denn ich bemühte mich, eines der schwierigsten Probleme in der IT zu lösen, nämlich »Dinge zu benennen«. Trotz meiner Befürchtung, harsche Kritik zu ernten, da ich mit frevelhaften Worten unsere Sichtweise auf Daten grundlegend veränderte, wurde der Vortrag vom Publikum positiv aufgenommen. Die Schmerzen von Menschen, die mit Daten arbeiten – Data Analysts oder Data Scientists – waren real; sie alle bemühten sich, zeitnah Zugriff auf qualitativ hochwertige und vertrauenswürdige Daten zu erhalten. Das Gleiche galt auch für die Data Engineers, die versuchen, Daten aus unzuverlässigen Datenquellen in eine Form zu bringen, die andere nutzen können, und das alles ohne engen Kontakt zur Fachabteilung. Die Führungskräfte im Publikum nickten bei der Feststellung, dass die Rendite ihrer Daten- und Analyselösungen nur mittelmäßig war. Ich verließ die Konferenz mit mehr Vertrauen in das, was nach den Lakes kommen könnte. Ein paar Monate später verpasste ich ein einwöchiges Treffen des Tech Advisory Board in China. Meine dreijährige Tochter hatte in der Nacht vor dem Abflug aus den USA Fieber bekommen. Ich schaffte es bis in das Flugzeug und verbarg meine Verzweiflung darüber, dass ich mich eine Woche lang von meinem kranken Kind trennen musste, aber als der Pilot der Besatzung ankündigte, dass die Türen des Flugzeugs geschlossen würden, brach ich zusammen. Ich verließ das Flugzeug. Jetzt hatte ich eine Woche Zeit, mich zurückzuziehen und die Gedanken und Erfahrungen mit Data Mesh in einem Artikel mit dem Titel »How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh« (<https://oreil.ly/rxjiW>) in Worte zu fassen, der freundlicherweise von Martin Fowler gehostet wurde. Der Artikel war ein voller Erfolg und wurde unglaublich oft gelesen, so als hätte ich gerade die Worte gesagt, an die andere im Stillen bereits gedacht hatten. Drei Jahre später geht nun dieses Buch detailliert darauf ein, warum Data Mesh wichtig ist, was es umfasst und wie man es umsetzt.

Warum ich dieses Buch geschrieben habe und warum gerade jetzt

In den wenigen Jahren, die seit der Vorstellung von Data Mesh vergangen sind, hat es enormen Anklang bei den Unternehmen gefunden, die es eingeführt hatten. Es hat Anbieter dazu ermutigt, zu versuchen, ihre Produkte so anzupassen, dass sie für Data-Mesh-Implementierungen geeignet sind. Es hat eine stetig wachsende Community geschaffen, die ihre Erfahrungen austauscht.

Trotz dieser rasanten Entwicklung schreibe ich dieses Buch vielleicht etwas früher, als ich es mir gewünscht hätte. Wir befinden uns noch in den Anfangsjahren eines grundlegend anderen Ansatzes bei der Bereitstellung und Erstellung von Daten für analytische Anwendungsfälle und Machine Learning. Aber unsere Branche hat die Tendenz, neue Konzepte und Buzzwords bis zur Unkenntlichkeit zu verdrehen.

Daher habe ich beschlossen, jetzt dieses Buch zu schreiben, um eine gemeinsame Grundlage für künftige Entwicklungen von Data-Mesh-Implementierungen zu schaffen. Ich wollte sicherstellen, dass wir, bevor wir uns dazu hinreißen lassen, neue technische Lösungen zu entwickeln, verstehen, warum wir etwas ändern müssen, welche Probleme wir lösen wollen und wie wir das tun sollten.

Dieses Buch schafft eine Grundlage für *die Ziele* von Data Mesh, warum wir uns damit beschäftigen sollten und für seine *Grundprinzipien*. Wir schauen uns an, wie man die Grundprinzipien anwendet, um eine High-Level-Architektur zu schaffen, und ich gebe Ihnen Werkzeuge an die Hand, mit denen Sie die Implementierung *umsetzen* und die *Organisation* und *Kultur* verändern können.

Wer dieses Buch lesen sollte

Dieses Buch richtet sich an Menschen mit den unterschiedlichsten Rollen und Kompetenzen. Data Mesh ist ein Paradigmenwechsel, und es erfordert den gemeinsamen Einsatz vieler sich ergänzender Rollen und Disziplinen in Bereichen wie Softwarearchitektur, Softwareentwicklung und Administration bis hin zum Produkt- und Top-Level-Management sowie den Führungskräften, um es für ein Unternehmen Wirklichkeit werden zu lassen.

Hier ist eine kurze Zusammenfassung der Personas der Leserinnen und Leser und was sie aus diesem Buch mitnehmen können:

- *Nutzer analytischer Daten* wie Data Scientists und Data Analysts sollten dieses Buch lesen, um zu verstehen, was Data Mesh ihnen ermöglicht. Sie lernen, wie sie ihrerseits ihre Erkenntnisse und Schlussfolgerungen als neue Datenprodukte im Data Mesh bereitstellen.
- *Datenlieferanten*, wie Entwicklungsteams oder Data Engineers, sollten dieses Buch lesen, um zu verstehen, wie Data Mesh die beiden Welten der operativen und analytischen Daten und Anwendungen zusammenbringt. Sie werden sehen, wie ihre Rollen in funktionsübergreifende Domänenteams übergehen und welche Art von Architektur sie aufbauen müssen, um Data Mesh zu ermöglichen.
- *Infrastruktur-Product-Owner, Softwarearchitekten und Softwareentwicklerinnen* sollten dieses Buch lesen, um die Rolle und das Design einer Self-Service-Datenplattform zu verstehen, um eine Reihe von gut integrierten Diensten bereitzustellen, die es funktionsübergreifenden Domänenteams ermöglichen, Daten dezentral im großen Umfang zu teilen.
- *Data-Governance-Teams* sollten dieses Buch lesen, um die neue Struktur und den neuen Ansatz zur Erreichung von Governance-Zielen zu verstehen, die eine unabhängige Domänenverantwortung für Daten fördern, organisatorische Engpässe beseitigen und sich stark auf Automatisierung stützen. Dieses Buch stellt eine neue Rolle und Form für Data Governance vor.

- *Führungskräfte und Manager* sollten dieses Buch lesen, um den bevorstehenden Paradigmenwechsel zu verstehen und zu lernen, eine auf Data Mesh basierende Datenstrategie zu formulieren, die Transformation durchzuführen und ihre Organisation auf diesem Weg zu begleiten.

Dieses Buch richtet sich sowohl an Personen, die sich mit Daten und deren Analysen befassen, als auch an diejenigen, die sich mehr auf die Entwicklung von Software und deren Betrieb konzentrieren. Data Mesh schließt die Lücke zwischen diesen beiden Gruppen.

Wenn Sie einen *Hintergrund in traditioneller Datenanalyse* haben, vielleicht als Data Engineer oder Data Analyst, möchte ich Sie ermutigen, Ihre Vorurteile aus der Vergangenheit abzulegen. Seien Sie offen für neue Wege, das Problem der analytischen Datenverwaltung und -verarbeitung zu lösen. Akzeptieren Sie die Automatisierung als unverzichtbaren Begleiter von Daten.

Wenn Sie einen *Hintergrund in Anwendungsentwicklung, Softwarearchitektur oder -infrastruktur* haben, lesen Sie dieses Buch mit Empathie für Daten und Analysen. Sehen Sie sich als Teil der Lösung für die Bereitstellung und die Nutzung von Daten zur Verbesserung Ihrer Anwendungen. Stellen Sie sich eine neue Zukunft vor, in der die Arbeit mit Daten und die Entwicklung von Anwendungen zwei sich ergänzende Elemente sind, um Ihre Lösungen erfolgreich zu machen.

Wie man dieses Buch liest

Ich empfehle Ihnen dringend, mit dem »Prolog: Fallstudie« zu beginnen. Dieses kurze Kapitel vermittelt Ihnen ein Gefühl dafür, wie Data Mesh funktioniert. Es veranschaulicht die Auswirkungen von Data Mesh in der täglichen Praxis. Es demonstriert die Prinzipien von Data Mesh anhand einer fiktiven Geschichte über ein digitales Streaming-Unternehmen: Daff, Inc.

Der Rest des Buchs ist in fünf Teile gegliedert:

Teil I, »Was ist Data Mesh?«

In diesem Teil werden die Grundprinzipien von Data Mesh vorgestellt und ihre transformativen Auswirkungen beschrieben. Ich hoffe, dass jeder diesen Teil des Buchs liest, da der Inhalt alle zukünftigen Diskussionen über Data Mesh prägen wird.

Teil II, »Warum Data Mesh?«

Wenn Sie sich nicht sicher sind, ob Data Mesh die richtige Wahl für Sie ist, oder wenn Sie verstehen wollen, welche Probleme es löst und wie es sie löst, oder wenn Sie einfach mitreden wollen, lesen Sie diesen Teil des Buchs. Er vergleicht Data Mesh mit der Vergangenheit und erörtert, warum das, was uns hierher gebracht hat, uns nicht in die Zukunft führen wird. Ich empfehle allen Leserinnen und Lesern, diesen Teil des Buchs sorgfältig zu lesen.

Teil III, »Wie entwirft man eine Data-Mesh-Architektur?«

Dieser Teil richtet sich an alle Menschen in der technischen Praxis und an Führungskräfte. Die Kapitel in diesem Teil des Buchs konzentrieren sich auf die High-Level-Architektur von Data Mesh und dessen Komponenten. Sie helfen Ihnen, Ihre Data-Mesh-Architektur zu entwerfen und Standardtechnologien auf ihre Eignung für Data Mesh zu prüfen.

Teil IV, »Wie entwirft man Datenprodukte?«

Dieser Teil befasst sich mit der detaillierten Gestaltung eines Kernkonzepts von Data Mesh, dem sogenannten Datenprodukt. Es soll komplexe Konzepte vereinfachen, ohne auf notwendige Details zu verzichten. Er sollte für alle Rollen in der Praxis und im Management verständlich sein. Personen, die in einer technischen Führungsposition bei der Implementierung verschiedener Aspekte von Data Mesh sind, sollten den größten Nutzen aus diesem Teil des Buchs ziehen.

Teil V, »Wie führt man Data Mesh ein?«

Dieser Teil ist ein Leitfaden für Personen in Funktionen, die Einfluss auf die gesamte Umsetzung der Datenstrategie und den dazugehörigen organisatorischen Wandel haben. Er enthält konkrete Ratschläge dazu, wie eine evolutionäre Einführung von Data Mesh angegangen werden kann und wie organisatorische Entscheidungen in Bezug auf Teamstruktur, Anreize, Kultur usw. getroffen werden können.

In diesem Buch verwendete Konventionen

In diesem Buch werden die folgenden typografischen Konventionen verwendet:

Kursiv

Kennzeichnet neue Begriffe, URLs, E-Mail-Adressen, Dateinamen und Dateierweiterungen.

Fett

Wird für Namen von Domänen und Datenprodukten verwendet.

Feste Zeichenbreite

Wird für Code sowie innerhalb von Absätzen verwendet, um auf Codeelemente wie Variablen- oder Funktionsnamen, Datenbanken, Datentypen, Umgebungsvariablen, Anweisungen und Schlüsselwörter hinzuweisen.

Feste Zeichenbreite in fett

Zeigt Befehle oder anderen Text an, der vom Benutzer genau so eingegeben werden muss.

Feste Zeichenbreite in kursiv

Zeigt Text an, der durch vom Benutzer eingegebene Werte oder durch kontextabhängige Werte ersetzt werden soll.



Dieses Element steht für einen allgemeinen Hinweis.



Dieses Element steht für eine Warnung oder einen Warnhinweis.

Danksagungen

Ich möchte dieses Buch meinem Partner, Adrian Paoletti, und meiner Tochter, Arianna Paoletti, widmen. Ohne ihre Geduld und selbstlose Liebe und Unterstützung würde es dieses Buch nicht geben. Wir haben in den letzten anderthalb Jahren viele gemeinsame Urlaube und Wochenenden verpasst, damit ich dieses Buch fertigstellen konnte. Ich werde ihnen für ihr Verständnis und ihre Liebe immer dankbar sein. Ich möchte dieses Buch auch meiner Mutter, Nayer Dadpay, und meiner Schwester, Parisa Dehghani, widmen, deren Liebe und ermutigende Worte mich behutsam durch den Prozess der Fertigstellung dieses Buchs geführt haben. Ich liebe euch alle.

Man sagt, ein Buch zu schreiben, sei ein einsames Unterfangen – nicht in meinem Fall. Ich möchte mich bei meinen aufmerksamen Rezensenten bedanken, die mich auf dem Weg des Schreibens begleitet und großzügig ihre Zeit und ihr Feedback zur Verfügung gestellt haben. In keiner besonderen Reihenfolge: Andy Petrella, vielen Dank für das Teilen deiner Perspektive als Data Scientist mit Bescheidenheit und Humor. Chris Ford, ich danke dir für deine durchdachten Kommentare zur Architektur, die meinen Blickwinkel oft erweitert haben. Mammand Zadeh, danke für die Stimme eines erfahrenen und doch pragmatischen Experten für Dateninfrastrukturen, der mir immer dabei hilft, den Bogen von der Idee zur Realität zu spannen. Martin Fowler, danke, dass du das große Ganze beleuchtest, mir die Lücken aufzeigst und mir hilfst, komplexe Konzepte zu veranschaulichen. Danilo Sato und Sam Ramji, ich danke euch für eure Ratschläge, eure Weisheit und eure Zeit.

Viele Mitarbeitende von Thoughtworks waren an der Entstehung wegweisender Themen in der Tech-Industrie beteiligt: Microservices, Continuous Delivery, Agilität, und die Liste geht weiter. Einer der Gründe dafür ist, dass die Führung von Thoughtworks genau die richtigen Bedingungen für die Verbreitung von Kreativität im Streben nach Software-Exzellenz schafft. Ich möchte mich bei Rebecca Parsons und Chris Murphy (keine besondere Reihenfolge) dafür bedanken, dass sie mich beim Schreiben dieses Buchs unterstützt haben. Ich möchte mich bei meinen ehemaligen und derzeitigen Kollegen bei Thoughtworks (ebenfalls ohne besondere Reihenfolge) bedanken: Gagan Madan, Zichuan Xiong, Neal Ford, Samia Rahman, Sina Jahangirizadeh, Ken Collier, Srikar Ayilavarapu, Sheroy Marker, Danilo Sato, Emily Gorcenski, David Colls, Erik Nagler und vielen anderen.

Ich möchte all den Menschen bei O'Reilly danken, die die Veröffentlichung dieses Buchs ermöglicht haben. Aus der wunderbaren und leidenschaftlichen Familie von O'Reilly möchte ich Gary O'Brien hervorheben. Gary, ich danke dir für deine unermüdliche Unterstützung und für all die Wochenenden, die du dir von deiner Familie freigenommen hast, um meine Inhalte zu prüfen und meine Fragen zu beantworten, um mich während der Tiefen und Zweifel aufzumuntern und mich wieder auf den richtigen Weg zu bringen. Melissa Duffield, danke, dass du dieses Buch auf den Weg gebracht hast, mir geholfen hast, den ersten Schritt zu tun, und mich mit deiner unbestreitbaren menschlichen Empathie unterstützt hast.

Schließlich möchte ich meinem Lehrer und Mentor während des Schreibens, Martin Fowler, danken. Martin, ich danke dir, dass du mich bei jedem Schritt unterstützt hast.

Danksagungen der Übersetzer

Zuallererst wollen wir uns bei Zhamak Dehghani bedanken, dass du mit Data Mesh den nächsten logischen Schritt in der Dezentralisierung der modernen Softwarearchitektur so klar beschrieben und damit auch eingeleitet hast. Dein Buch ist ein wundervoll illustriertes Standardwerk für die künftige Generation von Entwicklerinnen und Entwicklern. Zhamak, danke für deine wegweisenden Impulse!

Larysa Visengeriyeva, danke, dass du mit uns das Thema Data Mesh bei INNOQ aufbereitet hast – ohne dich wären wir nie so tief in das Thema abgetaucht und hätten auch nur halb so viel Spaß dabei gehabt. Und dir, lieber Stefan Tilkov, vielen Dank, dass du uns dies überhaupt ermöglicht hast.

Natürlich möchten wir uns auch besonders bei Alexandra Follenius von O'Reilly Deutschland bedanken für die freundliche und unkomplizierte Unterstützung. Danke, dass du unsere zahlreichen E-Mails immer schnell und ausführlich beantwortet hast.

Ein ganz besonderer Dank geht an unsere Familien, die uns an vielen Abenden und Wochenenden den Rücken freigehalten haben.

Prolog: Fallstudie

Die Vorstellungskraft trägt uns oft in Welten, die es nie gab. Aber ohne sie gehen wir nirgendwo hin.

– Carl Sagan

Auf jedes erfolgreiche Unternehmen kommen drei, die gescheitert und vergessen sind. Die Zahl der Gescheiterten übersteigt die Zahl der Überlebenden.¹ Im Zeitalter der künstlichen Intelligenz ist es kein seltsamer Zufall, dass die überlebenden und erfolgreichen Unternehmen die heutige Komplexität beherrschen und konsequent datenbasierte Experimente durchführen. Sie stellen sich dem ständigen Wandel und nutzen Machine Learning, um die Realität jenseits menschlicher Logik und Argumentation zu verstehen.

Daff, Inc.,² ein fiktives globales Musik- und Audio-Streaming-Unternehmen,³ ist ein Beispiel für ein solches Unternehmen. Daff hat seine Mission erfolgreich umgesetzt: »Kunstschaffende und Publikum auf der ganzen Welt miteinander zu verbinden in einem einzigartigen Musikgenuss in jedem Moment des Lebens.« Hinter der Mission von Daff stehen die hohen Erwartungen des Unternehmens an Daten, Analysen und künstliche Intelligenz, die durch einen Ansatz erreicht werden, den man als *Data Mesh* bezeichnet. Data Mesh ist die Grundlage der Datenstrategie, der Datenarchitektur und des Betriebsmodells von Daff. Es erlaubt dem Unternehmen, skalierbar und schnell mithilfe von Daten und *Machine Learning* (ML) zu experimentieren, zu lernen und sich anzupassen.

-
- 1 Nach Angaben (<https://oreil.ly/SrQSC>) des US Bureau of Labor Statistics (BLS) überleben nur 25 % der neu gegründeten Unternehmen 15 Jahre oder länger. Diese Statistiken haben sich im Laufe der Zeit nicht wesentlich verändert und sind seit den 1990er-Jahren ziemlich konstant.
 - 2 Der Name des fiktiven Unternehmens, »Daff«, stammt vom Namen eines persischen Schlaginstruments, das mehr als 3.000 Jahre alt ist und noch heute gespielt wird. Es dient als Symbol für etwas Langlebiges.
 - 3 Daff ist inspiriert von meinen Erfahrungen als Hörer von Spotify (<https://oreil.ly/rzLnp>). Die Beispiele für die interne Funktionsweise des Unternehmens und seiner Dienste sind alle frei erfunden.

Ich möchte Ihnen die Geschichte von Daff erzählen, nachdem es Data Mesh eingeführt hat. Anhand dieser Geschichte von Daff werden Sie das Wesentliche über Data Mesh erfahren. Sie werden sehen, wie die Prinzipien von Data Mesh angewendet werden. Sie lernen die Vorteile kennen und werden verstehen, wie die Architektur und die Organisationsstruktur aufgebaut sind und funktionieren.

Ich finde, ein komplexer Ansatz wie Data Mesh lässt sich am besten anhand eines Beispiels erklären. Die Entwicklung von Data Mesh steht jedoch noch zu sehr am Anfang, um ein echtes Beispiel für ein Unternehmen mit einem ausgereiften Data Mesh zu schildern. Schließlich sind wir gerade erst dabei, die ersten Data-Mesh-Umsetzungen zu entwickeln. Daher beschreibe ich hier ein fiktives Unternehmen, das die Merkmale aufweist, die ich in einigen Jahren erwarten würde. Auch wenn wir nicht davon ausgehen, dass die Realität genau unserer Vorstellung entsprechen wird, so ist unsere Vision von unserem Ziel doch ein wesentlicher Bestandteil des Verständnisses dessen, was wir zu erreichen versuchen. Um dieses Bild am besten zu vermitteln, schreibe ich über dieses fiktive Unternehmen so, wie es vermutlich in der Wirtschaftspresse präsentiert werden würde.

Während ich die Geschichte erzähle, werde ich Fußnoten setzen, damit Sie zu den späteren Kapiteln springen können, die tiefer in die Aspekte eintauchen, die ich hier nur kurz anspreche. Ich empfehle Ihnen, zunächst die ganze Geschichte zu lesen und erst nach deren Ende zu den späteren Kapiteln weiterzublättern.

Data Mesh in der Praxis

Wir schreiben das Jahr 2022.

Daff hat dank eines konsequenten Fokus auf die User Experience unter Einsatz von Machine Learning ein starkes Wachstum seiner Premium-Abonnenten verzeichnet. Das Unternehmen ist nach wie vor eine der beliebtesten Plattformen, die mithilfe von Daten ein einzigartiges Erlebnis schafft, eine umfangreiche Bibliothek von Inhalten kuratiert und neue und aufstrebende Künstlerinnen und Künstler erreicht. Daff hat sich kontinuierlich weiterentwickelt, indem neue Dienste hinzugefügt und in benachbarte Domänen wie Podcasts, Videos und die Organisation von Live-Events expandiert wurden. Heute ist Daff in fast jedem Land der Welt tätig und verfügt über ein wachsendes Netzwerk lokaler und globaler Geschäftspartner, von Event- und Kunstzentren bis hin zu Workout Plattformen.

In den letzten drei Jahren hat Daff die Verwaltung und Nutzung analytischer Daten auf einen Ansatz namens *Data Mesh* umgestellt. Data Mesh ist ein neuer, skalierbarer Ansatz zur Nutzung von analytischen Daten in großen Unternehmen, der Daten und Business enger als je zuvor zusammenbringt.

Daff setzt ausgefeilte Machine-Learning-Modelle ein, die kontinuierlich Muster in einem vielfältigen und sich ständig weiterentwickelnden Datenbestand innerhalb und außerhalb des Unternehmens auswerten. Das Unternehmen stellt seinen Hörerinnen und Hörern Empfehlungen bereit, die auf ihren Geschmack, ihre Stim-

mung, ihre Tageszeit und ihren Standort zugeschnitten sind. Mithilfe von Daten unterstützt Daff Künstlerinnen und Künstler mit gezielten Kampagnen, um deren Reichweite zu erhöhen. Mit Datenanalysen, Dashboards, Berichten und Visualisierungen hat Daff das Geschäft jederzeit und in Echtzeit im Blick. Wie Daff seine Daten nutzt, ist nur die Spitze des Eisbergs.

Schauen wir uns mal an, wie Daff das macht.

Eine Kultur der Neugierde und des Experimentierens mit Daten

Eine der bemerkenswertesten Veränderungen bei Daff ist die ausgeprägte Kultur, die es immer wagt zu fragen: »Was wäre, wenn ...«: Was wäre, wenn wir etwas ändern würden, um die Dinge nur ein kleines bisschen besser zu machen? Es gibt eine Kultur, die unablässig Experimente durchführt, die Ergebnisse beobachtet, die Daten analysiert, sie interpretiert, daraus lernt und sich anpasst.

Diese Kultur basiert auf einer technischen Grundlage, die es jedem leicht macht, Experimente zu wagen: große Experimente mit angewandtem Machine Learning oder kleine Experimente, bei denen einfach nur die Benutzeroberfläche optimiert wird.

Daff ist in Geschäftseinheiten organisiert, die als *Domänen* bezeichnet werden. Die **Player-Domäne** konzentriert sich auf den eigentlichen Musikplayer, der auf mobilen Geräten verwendet wird, die **Partnership-Domäne** setzt auf eine Zusammenarbeit in Geschäftsfeldern wie Fitnessanwendungen und Kunstveranstaltungen, und die **Playlist-Domäne** untersucht fortschrittliche Ansätze zur Erstellung von Wiedergabelisten. Jede Domäne kombiniert Softwareentwicklung und fachliche Kompetenzen und ist für die Softwarekomponenten zur Unterstützung der jeweiligen Domäne verantwortlich.

Wenn man sich bei Daff umschaut, stellt man fest, dass jede Domäne ständig zahlreiche Experimente durchführt, um ihre Anwendungen und Dienste zu verbessern. Zum Beispiel experimentieren die **Player-Teams** ständig mit einer besseren Interaktion mit den Nutzern. Das **Partnership-Team** experimentiert mit Daten, die von einer Vielzahl externer Quellen wie **Workout Platforms**, **Kunstzentren** usw. erfasst werden. Die **Playlist-Teams** wenden fortgeschrittenes Machine Learning an, um attraktive Musiksammlungen zu kuratieren und zu empfehlen. Und das **Artist-Team** setzt Machine Learning ein, um neue Künstlerinnen und Künstler zu entdecken, zu überzeugen und aufzunehmen, die normalerweise unentdeckt geblieben wären.

Alle fachlichen Domänen und die mit ihnen zusammenarbeitenden Technologie-teams wissen aussagekräftige, vertrauenswürdige und sichere Daten sehr zu schätzen. Und nicht nur das: Jeder erwartet, dass der direkte Zugriff auf Daten im gesamten Unternehmen die Norm ist. Alle wissen, welche Rolle sie dabei spielen, um dies zu ermöglichen. Sie sind alle für die Daten verantwortlich und interessieren sich auch dafür.

In jeder Domäne werden Machine-Learning-Modelle mit Begeisterung immer dann angewendet, wenn ein Feature oder eine Funktionalität durch die Auswertung historischer Daten und Muster umgesetzt werden kann. Die **Playlist-Teams** verwenden beispielsweise generative Machine-Learning-Modelle, um ausgefallene und wunderschöne Musiksammlungen zu erstellen. Die Sammlungen sind auf verschiedene Aktivitäten ausgerichtet, vom Laufen bis zum konzentrierten Lernen. Das **Artist-Team** nutzt mehrere Datensätze aus den sozialen Medien und von anderen Agenturen außerhalb von Daff, um aufstrebende Künstlerinnen und Künstler zu erkennen, zu fördern und mit ihrem neuen Publikum zu vernetzen.

Man spürt die Begeisterung für die Datennutzung und das Kennenlernen einer neuen Perspektive, in der man Signale entdecken kann, die für unsere menschlichen Sinne nur ein Rauschen wären.⁴

Die Datenkultur vor Data Mesh

Diese Kultur steht im krassen Gegensatz zu dem, wie Daff vor drei Jahren war. Datenerhebung, Experimente und Auswertungen waren an ein separates Datenteam ausgelagert worden. Das Datenteam stand unter enormem Druck. Die Domänen hatten kein Vertrauen in die Daten oder konnten die benötigten Daten oft nicht finden. Das Datenteam war ständig damit beschäftigt, die Probleme in der Daten-Pipeline zu lösen, die durch jede noch so kleine Änderung in den Upstream-Anwendungen und ihren Datenbanken verursacht wurden, oder die Anforderungen der ungeduldligen Fachbereiche zu erfüllen, die am besten noch gestern eine Lösung benötigten. Die Domänen selbst hatten keine Verantwortung und kein Interesse daran, dass Daten sowohl einfach verfügbar als auch zuverlässig und nutzbar waren. Die Dauer und die Hindernisse, um an die richtigen Daten zu gelangen, machten es für die Domänen unglaublich schwer, sich an neue Experimente zu wagen.

Der Vergleich der beiden Erfahrungen zeigt, wie weit Daff in den drei Jahren seit der Umstellung auf Data Mesh vorangekommen ist.

Eine enge Partnerschaft zwischen Daten und Machine Learning

Die Kultur des Experimentierens mit Daten scheint einfach zu schön, um wahr zu sein. Um zu sehen, wie sie in der Praxis aussieht, wollen wir die Geschichte eines datenorientierten Features nachvollziehen, an dem Daff kürzlich gearbeitet hat, und die Erfahrungen der beteiligten Personen beleuchten.

Smarte Wiedergabelisten sind ein erfolgreiches Feature der Daff-Plattform. Die **Music-Playlist-Domäne** hat an mehreren Machine-Learning-Modellen gearbeitet,

⁴ Siehe Kapitel 16, »Organisation und Unternehmenskultur«, für eine Diskussion über Werte, Kultur, Anreize und Verantwortlichkeiten in einer Data-Mesh-Organisation unter Verwendung des hier vorgestellten Beispiels Daff.