

Data Management at Scale, 2nd Edition

Table of contents

[Foreword](#)

[Preface](#)

[Why I Wrote This Book and Why Now](#)

[Who Is This Book For?](#)

[How to Read or Use This Book](#)

[Conventions Used in This Book](#)

[O'Reilly Online Learning](#)

[How to Contact Us](#)

[Acknowledgments](#)

[1. The Journey to Becoming Data-Driven](#)

[Recent Technology Developments and Industry Trends](#)

[Data Management](#)

[Analytics Is Fragmenting the Data Landscape](#)

[The Speed of Software Delivery Is Changing](#)

[The Cloud's Impact on Data Management Is Immeasurable](#)

[Privacy and Security Concerns Are a Top Priority](#)

[Operational and Analytical Systems Need to Be Integrated](#)

[Organizations Operate in Collaborative Ecosystems](#)

[Enterprises Are Saddled with Outdated Data Architectures](#)

[The Enterprise Data Warehouse: A Single Source of Truth](#)

[The Data Lake: A Centralized Repository for Structured and Unstructured Data](#)

[The Pain of Centralization](#)

[Defining a Data Strategy](#)

[Wrapping Up](#)

[2. Organizing Data Using Data Domains](#)

[Application Design Starting Points](#)

[Each Application Has a Data Store](#)

[Applications Are Always Unique](#)

[Golden Sources](#)

[The Data Integration Dilemma](#)

[Application Roles](#)

[Inspirations from Software Architecture](#)

[Data Domains](#)

[Domain-Driven Design](#)

[Business Architecture](#)

[Domain Characteristics](#)

[Principles for Distributed and Domain-Oriented Data Management](#)

[Design Principles for Data Domains](#)

[Best Practices for Data Providers](#)

[Domain Ownership Responsibilities](#)

[Transitioning Toward Distributed and Domain-Oriented Data Management](#)

[Wrapping Up](#)

[3. Mapping Domains to a Technology Architecture](#)

[Domain Topologies: Managing Problem Spaces](#)

[Fully Federated Domain Topology](#)

[Governed Domain Topology](#)

[Partially Federated Domain Topology](#)

[Value Chain–Aligned Domain Topology](#)

[Coarse-Grained Domain Topology](#)

[Coarse-Grained and Partially Governed Domain Topology](#)

[Centralized Domain Topology](#)

[Picking the Right Topology](#)

[Landing Zone Topologies: Managing Solution Spaces](#)

[Single Data Landing Zone](#)

[Source- and Consumer-Aligned Landing Zones](#)

[Hub Data Landing Zone](#)

[Multiple Data Landing Zones](#)

[Multiple Data Management Landing Zones](#)

[Practical Landing Zones Example](#)

[Wrapping Up](#)

[4. Data Product Management](#)

[What Are Data Products?](#)

[Problems with Combining Code, Data, Metadata, and Infrastructure](#)

[Data Products as Logical Entities](#)

[Data Product Design Patterns](#)

[What Is CQRS?](#)

[Read Replicas as Data Products](#)

[Design Principles for Data Products](#)

[Resource-Oriented Read-Optimized Design](#)

[Data Product Data Is Immutable](#)

[Using the Ubiquitous Language](#)

[Capture Directly from the Source](#)

[Clear Interoperability Standards](#)

[No Raw Data](#)

[Don't Conform to Consumers](#)

[Missing Values, Defaults, and Data Types](#)

[Semantic Consistency](#)

[Atomicity](#)

[Compatibility](#)

[Abstract Volatile Reference Data](#)

[New Data Means New Ownership](#)

[Data Security Patterns](#)

[Establish a Metamodel](#)

[Allow Self-Service](#)

[Cross-Domain Relationships](#)

[Enterprise Consistency](#)

[Historization, Redeliveries, and Overwrites](#)

[Business Capabilities with Multiple Owners](#)

[Operating Model](#)

[Data Product Architecture](#)

[High-Level Platform Design](#)

[Capabilities for Capturing and Onboarding Data](#)

[Data Quality](#)

[Data Historization](#)

[Solution Design](#)

[Real-World Example](#)

[Alignment with Storage Accounts](#)

[Alignment with Data Pipelines](#)

[Capabilities for Serving Data](#)

[Data Serving Services](#)

[File Manipulation Service](#)

[De-Identification Service](#)

[Distributed Orchestration](#)

[Intelligent Consumption Services](#)

[Direct Usage Considerations](#)

[Getting Started](#)

[Wrapping Up](#)

5. Services and API Management

[Introducing API Management](#)

[What Is Service-Oriented Architecture?](#)

[Enterprise Application Integration](#)

[Service Orchestration](#)

[Service Choreography](#)

[Public Services and Private Services](#)

[Service Models and Canonical Data Models](#)

[Parallels with Enterprise Data Warehousing Architecture](#)

[A Modern View of API Management](#)

[Federated Responsibility Model](#)

[API Gateway](#)

[API as a Product](#)

[Composite Services](#)

[API Contracts](#)

[API Discoverability](#)

[Microservices](#)

[Functions](#)

[Service Mesh](#)

[Microservice Domain Boundaries](#)

[Ecosystem Communication](#)

[Experience APIs](#)

[GraphQL](#)

[Backend for Frontend](#)

[Practical Example](#)

[Metadata Management](#)

[Read-Oriented APIs Serving Data Products](#)

[Wrapping Up](#)

[6. Event and Notification Management](#)

[Introduction to Events](#)

[Notifications Versus Carried State](#)

[The Asynchronous Communication Model](#)

[What Do Modern Event-Driven Architectures Look Like?](#)

[Message Queues](#)

[Event Brokers](#)

[Event Processing Styles](#)

[Event Producers](#)

[Event Consumers](#)

[Event Streaming Platforms](#)

[Governance Model](#)

[Event Stores as Data Product Stores](#)

[Event Stores as Application Backends](#)

[Streaming as the Operational Backbone](#)

[Guarantees and Consistency](#)

[Consistency Level](#)

[Processing Methods](#)

[Message Order](#)

[Dead Letter Queue](#)

[Streaming Interoperability](#)

[Governance and Self-Service](#)

[Wrapping Up](#)

[7. Connecting the Dots](#)

[Cross-Domain Interoperability](#)

[Quick Recap](#)

[Data Distribution Versus Application Integration](#)

[Data Distribution Patterns](#)

[Application Integration Patterns](#)

[Consistency and Discoverability](#)

[Inspiring, Motivating, and Guiding for Change](#)

[Setting Domain Boundaries](#)

[Exception Handling](#)

[Organizational Transformation](#)

[Team Topologies](#)

[Organizational Planning](#)

[Wrapping Up](#)

[8. Data Governance and Data Security](#)

[Data Governance](#)

[The Governance Framework](#)

[Processes: Data Governance Activities](#)

[Making Governance Effective and Pragmatic](#)

[Supporting Services for Data Governance](#)

[Data Contracts](#)

[Data Security](#)

[Current Siloed Approach](#)

[Trust Boundaries](#)

[Data Classifications and Labels](#)

[Data Usage Classifications](#)

[Unified Data Security](#)

[Identity Providers](#)

[Real-World Example](#)

[Typical Security Process Flow](#)

[Securing API-Based Architectures](#)

[Securing Event-Driven Architectures](#)

[Wrapping Up](#)

[9. Democratizing Data with Metadata](#)

[Metadata Management](#)

[The Enterprise Metadata Model](#)

[Practical Example of a Metamodel](#)

[Data Domains and Data Products](#)

[Data Models](#)

[Data Lineage](#)

[Other Metadata Areas](#)

[The Metalake Architecture](#)

[Role of the Catalog](#)

[Role of the Knowledge Graph](#)

[Wrapping Up](#)

[10. Modern Master Data Management](#)

[Master Data Management Styles](#)

[Data Integration](#)

[Designing a Master Data Management Solution](#)

[Domain-Oriented Master Data Management](#)

[Reference Data](#)

[Master Data](#)

[MDM and Data Quality as a Service](#)

[MDM and Data Curation](#)

[Knowledge Exchange](#)

[Integrated Views](#)

[Reusable Components and Integration Logic](#)

[Republishing Data Through Integration Hubs](#)

[Republishing Data Through Aggregates](#)

[Data Governance Recommendations](#)

[Wrapping Up](#)

[11. Turning Data into Value](#)

[The Challenges of Turning Data into Value](#)

[Domain Data Stores](#)

[Granularity of Consumer-Aligned Use Cases](#)

[DDSs Versus Data Products](#)

[Best Practices](#)

[Business Requirements](#)

[Target Audience and Operating Model](#)

[Nonfunctional Requirements](#)

[Data Pipelines and Data Models](#)

[Scoping the Role Your DDSs Play](#)

[Business Intelligence](#)

[Semantic Layers](#)

[Self-Service Tools and Data](#)

[Best Practices](#)

[Advanced Analytics \(MLOps\)](#)

[Initiating a Project](#)

[Experimentation and Tracking](#)

[Data Engineering](#)

[Model Operationalization](#)

[Exceptions](#)

[Wrapping Up](#)

[12. Putting Theory into Practice](#)

[A Brief Reflection on Your Data Journey](#)

[Centralized or Decentralized?](#)

[Making It Real](#)

[Opportunistic Phase: Set Strategic Direction](#)

[Transformation Phase: Lay Out the Foundation](#)

[Optimization Phase: Professionalize Your Capabilities](#)

[Data-Driven Culture](#)

[DataOps](#)

[Governance and Literacy](#)

[The Role of Enterprise Architects](#)

[Blueprints and Diagrams](#)

[Modern Skills](#)

[Control and Governance](#)

[Last Words](#)

[Index](#)

[About the Author](#)