Data Product Engineering for Agentic AI
Ramkumar Nottath, Nishchai JM

1. Tell us about yourself.

What are your qualifications for writing this book?

We bring a combined experience of over 30 years in data, analytics, and AI leadership roles across diverse industries. I (Nishchai J M) am a Senior Specialist Solutions Architect for Analytics at AWS, with deep expertise in modernizing data lakes, data warehouses, and scaling data platforms for AI/ML workloads. My work focuses on helping global enterprises build robust, future-ready data foundations. I've also been an active public speaker at global forums like AWS re:Invent, and an author for leading technical platforms such as the AWS Big Data Blog and IEEE Computer Society.

Ram Nottath, my co-author, is an accomplished tech leader for Data and Analytics at AWS who leads global initiatives in data and AI space, with a proven track record of driving data strategy and transformation for complex enterprises. His hands-on leadership spans from architecting modern data ecosystems to implementing AI-driven innovations for large-scale business impact. Ram's strategic viewpoint, paired with my technical depth, allows us to offer a 360-degree perspective — from boardroom strategy to implementation playbooks.

Together, our combined experience ensures that the book is not just theoretical but also packed with actionable insights, real-world case studies, and proven frameworks for navigating the unified data and AI journey.

Do you have any unique characteristics or experiences that will make you stand out as the author?

Yes, our partnership itself is a unique strength. We represent both ends of the data spectrum: Ram, as a visionary leader and business strategist, and myself, as a deeply technical architect and practitioner. This dual perspective is rare and enables us to address both executive concerns (like governance, ROI, and change management) and technical challenges (like data architecture, AI integration, and operationalization).

Additionally, we have firsthand experience guiding enterprises through the messy, real-world challenges of unifying siloed data environments and scaling AI initiatives responsibly. Our backgrounds span across industries such as travel & hospitality, financial services, retail, and manufacturing.

We're not just writing about future possibilities, we've lived them, advised C-level stakeholders, and built solutions at scale. Readers will get a candid, practitioner's view of what works, what doesn't, and what lies ahead in the evolving landscape of data and AI.

2. Tell us about the book's topic.

What is the technology or idea that you're writing about?

A data product is a reusable and maintainable data asset that provides consistent, reliable, and valuable information to serve specific business needs. It combines data, code, and infrastructure to deliver actionable insights or capabilities that can be consumed by various stakeholders. The

key aspects of data products include being self-contained and modular, well-documented and maintainable, providing clear value proposition, having defined quality metrics, and supporting scalability and reusability.

This book explores the critical intersection of data products and artificial intelligence, specifically focusing on the emerging paradigm of Agentic AI - autonomous AI systems that can plan, reason, and act independently toward specified goals. While numerous books cover machine learning and data engineering separately, this is the first comprehensive guide that will address how to design, build, and govern data products specifically for autonomous AI systems.

The technology focus is threefold: First, it covers the evolution from traditional machine learning to modern Agentic AI systems, including the architectural patterns and infrastructure needed to support autonomous decision-making. Second, it dives deep into data product engineering - showing how to build data systems that go beyond static storage to become intelligent, context-aware services that can support real-time agent interactions. Third, it addresses the critical governance and risk management frameworks needed when AI systems operate autonomously.

The book introduces novel concepts like agent-ready data products, which combine heterogeneous storage systems, semantic layers for contextual understanding, and embedded governance controls. It explores how traditional data architectures must evolve to support multi-modal data (text, images, events), real-time streaming, and dynamic feedback loops required by autonomous systems.

What makes this topic particularly timely is the rapid advancement of AI capabilities and Gartner's projection that by 2028, a significant portion of enterprise software applications will incorporate agentic AI. Organizations need practical guidance on building the data foundations that will enable safe and effective autonomous systems. This book bridges that gap by providing concrete frameworks, architectural patterns, and governance approaches specifically designed for the age of autonomous AI.

Why is it important now?

The emergence of autonomous, self-directing AI systems marks a critical shift from traditional AI, making this book essential now. With Gartner predicting widespread adoption of agentic AI by 2028, organizations urgently need guidance on building data foundations that can support autonomous decision-making. Traditional data management approaches are proving insufficient for AI systems that require real-time, contextual data access. This book provides timely, practical frameworks for organizations to bridge the gap between current data practices and the sophisticated requirements of autonomous AI systems, while addressing crucial governance and risk concerns.

In a couple sentences, tell us roughly how it works or what makes it different from its alternatives?

Unlike traditional books that focus solely on either data engineering or AI systems, this book uniquely bridges the gap by showing how to build and govern data products specifically designed for autonomous AI systems. It provides a comprehensive framework that combines modern data

architecture patterns with practical governance approaches, enabling organizations to create data products that can safely and effectively support AI agents' real-time decision-making capabilities. What sets it apart is its focus on the complete lifecycle of agent-ready data products, from architectural design to operational governance, addressing the distinct challenges of building data systems that can support autonomous, goal-seeking AI behavior.

3. Tell us about the book you plan to write.

What will the reader be able to do after reading this book?

1. Architect data foundations that enable AI agents to access, understand, and act upon diverse data sources in real-time, incorporating modern patterns like RAG pipelines, knowledge graphs, and semantic layers.

2. Build robust data products that go beyond traditional static datasets, featuring built-in governance controls, context-awareness, and feedback mechanisms essential for autonomous decision-making.

3. Implement comprehensive governance frameworks that address the unique challenges of autonomous AI systems, including dynamic decision audits, goal alignment, and risk management strategies.

4. Design organizational structures and workflows that support the development and maintenance of agent-ready data products, including new roles like AI-focused data product managers.

5. Deploy practical safeguards and monitoring systems to ensure AI agents operate within defined parameters while maintaining operational trust and regulatory compliance. The book aims to enable readers to master the complete spectrum of modern AI development while following best practices for performance, cost optimization, and responsible AI development.

Is your book designed to teach a topic or to be used as a reference?

This book is primarily designed as a comprehensive learning guide that follows a logical progression, but it's structured to also serve as an effective reference resource.

As a learning guide, it follows a carefully sequenced journey - starting with the evolution from traditional ML to Agentic AI, then building up through data product fundamentals, advanced architectural patterns, and governance frameworks. Each chapter builds upon previous concepts, helping readers develop a complete understanding of how to build and manage data products for autonomous AI systems.

As a reference, each chapter is self-contained with clear frameworks, architectural patterns, and practical implementations. The modular organization allows practitioners to quickly locate specific guidance on topics like:

- Architectural patterns for agent-ready data products

- Governance frameworks for autonomous systems
- Implementation guides for semantic layers and RAG pipelines
- Risk management strategies and controls
- Organizational design patterns

The book includes practical appendices, templates, and checklists that serve as quick reference materials for ongoing projects. While readers will benefit most from reading the book sequentially first, they can later return to specific sections as needed during implementation phases of their projects.

Does this book fall into a Manning series such as In Action, In Practice, Month of Lunches, or Grokking?

We are not sure on this. Can you please provide more context?
Are there any unique characteristics of the proposed book, such as a distinctive visual style, questions and exercises, supplementary online materials like video, etc?

4. Q&A

What are the three or four most commonly-asked questions about this technology?

1. How do we evolve our existing data infrastructure to support autonomous AI systems?
2. How can we ensure autonomous AI systems make reliable and governed decisions?
3. What's the right organizational structure and skill set needed to build and maintain agent-ready data products?
4. How do we manage risk and ensure compliance when deploying autonomous AI systems?

5. Tell us about your readers.

Your book will teach your readers how to accomplish the objectives you've established for the book. It's critical to be clear about the minimum qualifications you're assuming of your reader and what you'll need to teach them.

What skills do you expect the minimally-qualified reader to already have? Be specific.

1. Technical Background:
   - Basic understanding of different types of data storage systems - databases, data warehouses, datalakes
   - Fundamental knowledge of AI and machine learning concepts
   - Understanding of API concepts and data integration
   - Basic understanding of data pipelines and workflows
   - Knowledge of data governance concepts

2. System Architecture:

- Basic understanding of system architecture principles
- Familiarity with real-time and batch processing concepts
- Understanding of basic infrastructure components

3. Business Context:
    - Understanding of product management concepts
    - Basic knowledge of business requirements and SLAs
    - Awareness of compliance and regulatory requirements

What are the typical job roles for the primary reader? Be specific:

1. Data Engineers
    - Responsible for building and maintaining data infrastructure
    - Working with data pipelines and integration systems

2. AI/ML Engineers
    - Developing and implementing AI agent systems
    - Working on machine learning models and autonomous systems

3. System Architects
    - Designing data product architectures
    - Planning infrastructure for AI systems

4. Technical Product Managers
    - Managing data products
    - Overseeing AI system implementations

5. Data Product Managers
    - Specifically focused on data product development
    - Managing data-driven solutions

6. Solutions Architects
    - Designing end-to-end solutions for AI systems
    - Integrating data products with existing infrastructure

7. DevOps Engineers
    - Managing operational aspects of data products
    - Implementing monitoring and maintenance systems

What will motivate the reader to learn this topic?

1. Industry Transformation
    - The evolution of data products from passive to autonomous systems

- The growing importance of AI agents in business operations
- The need to stay current with emerging technologies

2. Career Growth
    - Growing demand for professionals who understand agentic AI systems
    - Emerging opportunities in autonomous data product development
    - Increasing need for expertise in AI-driven data infrastructure

3. Technical Advancement
    - Understanding how to build AI-ready data products
    - Learning to implement advanced monitoring and governance systems
    - Mastering real-time data processing and integration

4. Business Impact
    - Ability to create more efficient and autonomous systems
    - Skills to implement self-improving data products
    - Knowledge to reduce operational costs through automation

6. Tell us about the competition and the ecosystem.

What are the best books available on this topic and how does the proposed book compare to them?

The competition in this topics / ecosystem are:

1. Building Agentic AI Systems - This book focuses primarily on creating autonomous AI agents and covers planning and adaptation mechanisms. The book has limited coverage of data product architecture and governance.
2. AI Agents in Action - This book concentrates on implementing AI agents and their behavioural aspects. The book details out on agent development. While it covers agent architecture, it doesn't address the data product aspect.
3. Building AI-Powered Products - The book provides coverage on AI and generative AI product management. But predates the specific challenges of agentic AI systems.
4. Data Mesh: Delivering Data-Driven Value at Scale - This is one of the first books published to introduce "Data as a Product". While this is strong on data architecture, it doesn't address the agentic AI requirements and newer patterns because this was written before the emergence of agentic AI as a major force.

The proposed book fills the gap in the current ecosystem by specifically addressing how to design and manage data products that can effectively support autonomous AI agents, making it a valuable addition to the existing collection of books in this space.

What are the best videos available on this topic and how does the proposed book compare to them?

https://www.youtube.com/watch?v=x4OwZnsQ5Ng : An end-to-end data strategy is a unified and integrated blueprint for analytics and generative AI applications to manage, access, and analyze all your data. AWS provides a comprehensive set of services to store, share, query, and analyze data to easily integrate across data lakes and data warehouses with high-performance SQL and Apache Spark analytics. AWS gives you the tools to catalog and govern that data through business intelligence, machine learning, and generative AI.

https://www.youtube.com/watch?v=KrRD7r7y7NY : In recent years, the spotlight in AI has primarily been on large language models (LLMs) and emerging large multi-modal models (LMMs). Now, building on these tools, a new paradigm is emerging with the rise of AI agents and agentic reasoning, which are proving to be both cost-effective and powerful for building numerous new applications. As AI continues to evolve, data across all industries—particularly unstructured data such as text, images, video, and audio—is becoming more critical than ever. In this keynote session from BUILD 2024, Andrews Ng, Founder and Executive Chairman of Landing AI, explores the rise of AI, agents, and the growing role of unstructured data. He also discusses how this convergence will shape automation and application building across industries.

What other resources would you recommend to someone wanting to learn this subject?

1. https://medium.com/@community_md101/how-ai-agents-data-products-work-together-to-support-cross-domain-queries-decisions-for-3129b1d58c5e
2. https://www.decube.io/post/agentic-ai-data-governance
3. https://www.xenonstack.com/blog/agentic-ai-data-quality
4. https://www.techtarget.com/searchdatamanagement/opinion/Generative-AI-shines-spotlight-on-data-governance-and-trust
5. https://weareoakland.com/blog/role-of-data-governance-in-ai/
6. https://www.linkedin.com/pulse/critical-link-between-data-quality-generative-ai-eric-roch-w0krc

What are the most important web sites and companies?

1. **AWS** – aws.amazon.com

- Services: S3, Glue, Lake Formation, Redshift, DataZone, Bedrock, SageMaker, Amazon Q
- Leader in cloud-based data lakes and AI infrastructure.

2. **Google Cloud –** cloud.google.com

- Services: BigQuery, Vertex AI, Dataplex
- Strong in analytics, AI, and data management.

3. **Microsoft Azure** – azure.microsoft.com

- Services: Synapse Analytics, Purview, Fabric, OpenAI integration

Where do others interested in this topic gather?

People interested in unified data for AI gather at major conferences like AWS re:Invent, Google Cloud Next, Snowflake Summit, and the Data + AI Summit by Databricks, where industry leaders discuss trends in data lakes, governance, and AI-driven analytics.

Online communities such as LinkedIn Groups (Big Data & AI), Reddit (r/dataengineering, r/datascience), and Slack/Discord channels for data engineers provide spaces for discussions, knowledge sharing, and networking.

Research and industry organizations like LF AI & Data Foundation, Apache Foundation, and IEEE/ACM AI groups also play a crucial role in advancing unified data strategies through open-source projects, technical papers, and professional networking opportunities.

7. Book size and illustrations

Please estimate:

The approximate number of published pages to within a 50-page range

300 - 350

The approximate number of diagrams and other graphics

250

The approximate number of code listings

50

8. Contact information

Name: Ramkumar Nottath
LinkedIn: https://www.linkedin.com/in/ramnottath/

Name Nishchai JM,
Website/blog: https://www.theedgereview.org/magazine/volume-2-issue-1/how-dynamic-data-governance-enables-instant-insights-for-unprecedented-innovation
Linkendin: https://www.linkedin.com/in/nishchai-j-m-558b5032/

9. Schedule

To write and revise a chapter, most authors require 2-4 weeks. Please estimate your writing schedule

Chapter 1: We are assuming 4 – 6 weeks for Chapter 1 considering initial reviews and incorporating them.

- 1/2 manuscript: June 30th
- 2/2 manuscript: July 15th

Rest of Part 1 Chapter 2 and 3

- Manuscript: Aug 30th

Part 2 Chapter 4 – 6

- Manuscript: Oct 15th

Part 3 Chapter 7 – 9

- Manuscript: Nov 15th

Are there any critical deadlines for the completion of this book? New software versions? Known competition? Technical conferences?

10. Table of Contents

Part 1: Foundations for Data Products for Agentic AI

1. **Understanding Data Products for Agentic AI**
    - 1.1. What is a Data Product?
    - 1.2. Key Traits: Discoverability, Reusability, Observability, Versioning
    - 1.3. What is Agentic AI?
    - 1.4. What Are AI Agents? From Reactive Bots to Autonomous Planners
    - 1.5. How Data Products Power Autonomy, Context, and Feedback
    - 1.6. Summary

This chapter will introduce the core concepts of data products and their critical role in modern autonomous AI systems. It begins by defining what is a data product and its key traits such as discoverability, observability and versioning. It then moves on to explain the concept of agentic AI and AI agents, tracing their evolution from simple reactive bots to sophisticated autonomous planners. The chapter highlights why successful agentic AI applications require products designed specifically for agentic AI systems and emphasizes the need for real-time data, context awareness, and continuous feedback. The chapter also explores how a well-designed data product helps with autonomy, adaptability, and intelligent decision making.

2. **Data Demands for Agentic AI**
    - 2.1. The Nature of Agentic Behavior: Goal-Driven, Adaptive, Reflective
    - 2.2. Data challenges in Agentic AI
    - 2.3. How Agentic AI Consumes and Depends on Data
      Characteristics of AI Ready Data products
    - 2.4. Bridging Agent Architecture with Data Architecture
    - 2.5. Summary

This chapter explores the fundamental nature of agentic behavior, focusing on how these systems exhibit goal-driven decision-making, adaptive responses, and reflective capabilities. The chapter then moves into the various data challenges in agentic AI and unique ways that agentic AI systems consume and depend on data, establishing why traditional data management approaches are insufficient for these advanced systems. The chapter will also focus on how these agents process and utilize data differently from traditional AI systems, highlighting the need for more sophisticated data architectures.

3. Components of an Agentic AI–Ready Data Product

3.1.Heterogeneous Storage for Multi-Modal Data

3.2.Metadata and Semantic Enrichment

3.3.Built-In Observability and Usage Telemetry

3.4.Explainability Hooks for Agent Decisions

3.5.Resilience and Error Handling

3.6.Summary

This chapter outlines the key components of an Agentic AI-Ready Data Product. The system is built on Heterogeneous Storage that handles multiple data types, enhanced by Metadata and Semantic Enrichment for contextual understanding. It includes Built-In Observability for usage monitoring, Explainability Hooks to track AI decision-making, and Resilience and Error Handling to ensure system stability. These elements work together to create a reliable and transparent data infrastructure supporting AI operations.

Part 2: Data Architecture and Infrastructure for Agentic AI systems

4. Architecting Agentic AI-Ready Data Products

4.1.Real-Time Responsiveness: From Batch to Streaming

4.2.Supporting Multimodal Inputs: Text, Vision, Event Streams

4.3.Building Feedback-Integrated Pipelines

4.4.Aligning Schema Evolution with Agent Learning

4.5.MCP servers for Agent-Product Interactions

4.6.Example Architectures and Patterns in Action

4.7.Summary

This chapter dives into the architectural foundations necessary for creating data products that effectively support agentic AI systems. The chapter begins by exploring how data products must evolve from traditional batch processing to real-time streaming capabilities, enabling immediate responsiveness to agent needs. It then examines the complexities of handling multimodal inputs, showing how data products must seamlessly integrate text, visual data, and event streams to support comprehensive agent understanding.

The chapter demonstrates how feedback-integrated pipelines can be constructed to enable continuous learning and adaptation, while explaining the critical relationship between schema evolution and agent learning processes. The discussion extends to the design of specialized APIs, MCP servers and interfaces that facilitate efficient agent-product interactions, ensuring smooth communication between autonomous systems and data resources. The chapter concludes with practical implementations, showcasing example architectures and patterns in action.

5. Infrastructure to Support Agentic Data Products

- 5.1.Advanced Data Integration
- 5.2.Heterogeneous Data stores
  - 5.2.1.Traditional data stores - Data Lakes, Warehouses, and Lakehouses
  - 5.2.2.Vector Databases and Embedding Stores
- 5.3.Context aware data retrieval
  - 5.3.1.Retrieval-Augmented Generation (RAG) and Context Injection
  - 5.3.2.Knowledge Graphs and Semantic Data Layers
- 5.4.Real-Time Orchestration Tools and Streaming Frameworks
- 5.5.Summary

This chapter explores the foundational infrastructure needed for building intelligent data products, starting with advanced data integration across heterogeneous data stores and traditional data platforms. It examines how vector databases and embedding stores enable sophisticated context-aware retrieval, while demonstrating the critical role of RAG pipelines and context injection in enhancing agent reasoning capabilities. The chapter walks through how knowledge graphs and semantic layers provide structured understanding of relationships supported by real-time orchestration tools for dynamic data processing.

6. Governing Data Products in Agentic Ecosystems

- 6.1.Why Traditional Governance Fails in Autonomous Systems
- 6.2.Embedding Governance Metadata in Data Products
- 6.3.Ensuring Lineage and Policy Enforcement
- 6.4.Monitoring for Drift, Bias, and Misalignment
- 6.5.Auditability, Explainability, and Regulatory Readiness
- 6.6.Summary

This chapter addresses data product governance in autonomous systems. It explains why traditional governance methods are inadequate, emphasizes the importance of embedded governance metadata, and outlines key aspects including lineage tracking, policy enforcement, and monitoring systems for drift and bias. The chapter also covers auditability and regulatory compliance requirements, creating a comprehensive framework for managing data products in autonomous environments.

Part 3: Operations

7. **Risks and Mitigations in Agentic AI systems**
    - 7.1. Data Drift in Live Agent Systems
    - 7.2. Simulation Testing and Chaos Engineering for Agents
    - 7.3. Real-Time Anomaly Detection and Alerting
    - 7.4. Designing Human-in-the-Loop Safeguards
    - 7.5. Fallback Protocols and Disaster Recovery
    - 7.6. Summary

This chapter covers key risk management and mitigation strategies for AI agent systems. It covers data drift in live systems and how to manage performance degradation over time. The chapter explores simulation testing and chaos engineering approaches to identify potential failures, alongside real-time anomaly detection and alerting mechanisms. It discusses the integration of human oversight through well-designed safeguards and concludes with essential fallback protocols and disaster recovery procedures. These elements combine to create a robust framework for maintaining safe and effective AI agent operations.

8. **Operationalizing Data Products for Agentic Systems**
    - 8.1. Product Management Mindset for Data
    - 8.2. Data Contracts and SLAs for Agents
    - 8.3. In-built Governance in Data Interfaces
    - 8.4. Monitoring Agent-Data Interactions
    - 8.5. Case Study: (Autonomous Customer Support Systems - This could change)
    - 8.6. Case Study: (Industrial AI Agents in Manufacturing Pipelines - This could change)
    - 8.7. Summary

This chapter covers the operationalization of data products for agentic systems, covering key areas including product management approaches to data, data contracts and SLAs for agents, practice of embedding governance principles directly into data product interfaces, and monitoring of agent-data interactions. The chapter emphasizes treating data as a strategic asset and includes case studies on autonomous customer support and industrial AI manufacturing applications (the case studies may change). The focus is on establishing effective frameworks for managing data within agentic AI operations while ensuring quality, compliance, and reliability.

9. **Emerging Techniques in Autonomous Data Products**
    - 9.1. Current Developments in Autonomous Data Systems
    - 9.2. Experiments on Self-Monitoring and Regulatory Features
    - 9.3. Summary

This chapter explores current experimental techniques and developing practices in autonomous data products and ongoing research. The chapter discusses current researches in the areas such as intelligent pipelines with feedback mechanisms from AI agents, self-monitoring implementations, and parameter adjustment systems.