Data Modeling in the Age of Big Data

Training Details

Training Time	:	1 Day
Capacity	:	10
Prerequisites	:	There are no prerequisites for this course.

About Training

About Training

Data modeling is still an important process—perhaps more important than ever before. But data modeling purpose and processes must change to keep pace with the rapidly evolving world of data. This course examines the principles, practices, and techniques that are needed for effective modeling in the age of big data.

What You'll Learn

- To distinguish between data store modeling (schema on write) and data access modeling (schema on read) and when each is useful
- The elemental characteristics of data that provide a common denominator for data modeling for all types of data
- How the common denominator is used to map various kinds of databases, including relational, dimensional, NoSQL, NewSQL, graph, and document
- When traditional logical-to-physical modeling works and when it makes sense to reverse the process as physical-to-logical
- Trade-offs between methodological rigor and discovery-driven exploration in data modeling

Who Should Attend

Data architects; data modelers; database developers; data integrators; data analysts; report developers; anyone else challenged with the need to make structured enterprise data and non-traditional data sources work together.

Outline

Module 1 – Big Data Fundamentals

What is Big Data

- Big Data
- NoSQL
- Structured Data
- Beyond Structured Data

Big Data Opportunities

- Beyond Enterprise Data
- Beyond Transactions
- Understanding Cause and Effect
- Business Impact

NoSQL Technologies

- Relational Technology
- Key-Value Stores
- Document-Oriented Databases
- Graph Databases
- Summary of Database Technologies
- Vendor Landscape

Big Data Challenges

- Beyond Enterprise Data
- Multiple Management Platforms
- Lack of Fixed Schema
- Multiple Uses for Data
- Traditional Focus on Transactions
- Relational Perspective

Exercise: Big Data Opportunities

Module 2 – Modeling and Data Models

Models

- What is a Model?
- What is a Data Model?
- Why Model Data?
- More than a Diagram

Modeling for Relational Storage

- Relational Storage and BI
- Fixed Structure and Content
- Schema on Write
- Requirements First
- Data Modelers and Architects

Modeling for Non-Relational Storage

- Big Data and BI
- Flexible Schema
- Big Data Notation
- Schema on Read
- Data First, Requirements Last
- Business SMEs, Analytic Modelers, and Programmers

Complementary Approaches

- Relational and Non-Relational Data
- Incremental Value of Big Data
- Rigor vs. Agility
- Roles

Exercise: Modeling Purpose

Module 3 – Key-Value Stores

Key-Value Stores Defined

- The Basics
- NoSQL Foundation

Key-Value Data Representation

- Representing Things
- Representing Identities
- Representing Properties
- Representing Associations
- Representing Metrics

Use Cases

- Embedded Systems
- High-Performance In-Process Databases
- NoSQL Foundation

Examples

- Common Key-Value Store Products

Exercise: Key-Value Pairs Modeling

Module 4 – Document Stores

Document Stores Defined

- Document-Oriented Databases
- Basic Terminology
- Flexible Internal Structure
- Document Stores and Key-Value Stores
- Fields Can Have Multiple Values
- Fields Can Contain Sub-Documents
- Summary of Characteristics

Document Data Representation

- Representing Things
- Representing Identifiers

- Representing Properties
- Representing Associations
- Representing Metrics

Use Cases

- Choosing Document Storage
- Capture: Data Arrives in Document Format
- Explore Sources that Track Information Differently
- Augment
- Extend

Examples

- Common Document Store Databases

Exercise: Document Modeling

Module 5 – Graph Databases

Graph Databases Defined

- The Basics
- Data about Relationships
- The Terminology – Nodes and Edges
- The Terminology – Hyperedges
- The Terminology – Properties

Graph Data Representation

- Representing Things
- Representing Identities
- Representing Associations
- Representing Properties
- Representing Metrics

Use Cases

- Social Networks
- Network Analysis and Visualization
- Semantic Networks

Examples

- Common Graph Database Products

Module 6 – Embracing Big Data

BI Programs and Big Data

- Big Data and Information Asset Management
- The Gaps
- What Is Lost with Non-Relational
- BI and Analytics Gap
- Role/Skill Gaps
- Organization and Planning
- Balancing Standards with Flexibility
- Organize Around Purpose, Not Tools

- IAM Roadmap Including Big Data
- Architecture Still Important
- The Journey
- Cataloging and Prioritizing Opportunities
- Evolving Skills
- Technology Decision Models
- Responding to Tool Failures

Human Side of Big Data

- Changing Role of Data Modeling
- Traditional Data Modeler Role
- More Roles Doing Data Modeling
- When Data Modeling Occurs
- Merging Data Modeling and Profiling

Tapping Into Big Data

- Process Agility and Flexibility Over Formality
- More Exploration, Iteration, and Risk
- Importance of Metadata

Taking the Next Steps

- Conversations to Gather Opportunities
- Proofs of Concept
- Business Case / ROI
- Ongoing Value of Data Modeling
- New Tools, Same Workbench

Exercise: Embracing Big Data

Module 7 – Summary and Conclusion

Summary of Key Points

- A Quick Review

References and Resources

- To Learn More