



Introduction to NoSQL

by William McKnight



Module 0. About the Course (8 min)

Module 1. Big Data Overview and Common Themes (49 min)

- Overview
- No More One Size Fits All
- The No Reference Architecture
- The Relational Database Data Page
- What Does Big Data Mean?
- Google Search Trends
- Why the Sudden Explosion of Interest
- What Happens in an Internet Minute?
- Sensors Data
- Customer Demands Drive Technology
- New Data Types
- Benefits of JSON
- Why NoSQL for Big Data?
- ACID
- Hadoop, MapReduce and Big Data
- Why NoSQL Not Hadoop for Operations
- MapReduce Part 1 & 2
- Scale Up vs. Scale Out
- DFS Block Placement Example
- File System Summary

Module 2. NoSQL History and Jargon (17 min)

- Overview
- NoSQL Inspirations
- NoSQL History
- Google MapReducer Paper
- Google Bigtable Paper
- Memcached
- Schemaless
- Keeping it Simple
- CAP Theorem Part 1 & 2
- Automatic Sharding
- NoSQL Node Specification

Module 3. Enablers for NoSQL (19 min)

- Overview
- Data Integration
- Data Visualization
- Infrastructure Strategy, Including Cloud
- Traditional Data Modeling
- Data Modeling for NoSQL
- NoSQL is for Applications, Not DW or ERP
- NoSQL Schemaless Data Modeling
- Force Fitting Unstructured Data into RDBMS
- NoSQL Modeling from RDBMS



BD-02: Introduction to NoSQL

- *Security Concerns*
- *Easing Into Change*
- *What Will Motivate IT to Adopt NoSQL?*

Module 4. NoSQL Data Models (49 min)

- *Overview*
- *Data Types and NoSQL Data Models*
- *Key Value Stores*
 - Key Value Stores
 - Technical Characteristics
 - Key Value Stores Are Good For...
 - Leading Vendors
 - Mapping RDBMS Knowledge
 - Key Value Stores: Poor Uses
- *Document Stores*
 - Document Oriented Databases
 - Technical Characteristics
 - Document Oriented Databases are Good For...
 - Leading Vendors
 - JSON Document Example
 - Mapping RDBMS Knowledge
 - Detailed Look at Couchbase
 - Couchbase Server
 - Key Capabilities
 - Couchbase Data
 - Couchbase Can Be Either...
 - Storing and Retrieving Documents
 - Auto Sharding
 - Write Operations
 - Multit-Node Operations
 - Sample Operations
 - Optimistic Concurrency with CAS
- *Column Stores*
 - Column Stores Are Good For...
 - Leading Vendors
 - Column Stores Example
 - Mapping RDBMS Knowledge
 - Column Stores Poor Uses
 - Cassandra Background
 - Detailed Look at Cassandra
 - Cassandra Insert and Delete
 - Use Cases
 - Run Options
 - Quick Look at HBase
- *Operational Big Data Platform Solution*
- *Multiple NoSQL Solutions Working Together*

Module 5. Relationship Model: Graph Oriented (68 min)

- *Overview*
- *The Graph Database Revolution*



BD-02: Introduction to NoSQL

- *Relationship Data*
 - Realizing Value from Data Relationships in Consumer Web
 - High Business Value in Data Relationships
 - Unlocking Value from your Data Relationships
 - Relational DB's Can't Handle Data Relationships Well
 - Other SQL Databases Don't Handle Data Relationships
 - Use the Right Database for the Right Job
 - Value From Data Relationships: Common Use Cases
 - What Can be Vertices? Find the Network
 - What Can be Edges?
- *Graph Algorithms*
 - Page Rank
 - Page Rank: After First Results
 - Page Rank: Start of 2nd Iteration
 - Page Rank: After 2nd Iteration
 - Page Rank: After 3rd Iteration
 - Page Rank Iterations
 - Page Rank: 20 Iterations Until Convergence
 - Betweenness
 - Closeness
 - Eigen Centrality
 - Clustering Coefficient
 - Loopy Belief Propagation
- *Use Cases*
 - Great Questions for Graph Databases
 - The Ripple Effect
 - Bid Pricing
 - Pricing Competitively
 - Healthcare Fraud
 - The Small World Network Model
 - Resource Flow
 - Cargo/Traffic Management
 - Optimizing Transit Duration
 - Social Network Analysis
 - Telecommunications
 - Financial Fraudulent Detection
 - Plant Science
- *Graph Modeling*
 - Re-Imagine Your Data as a Graph
 - The Whiteboard Model is the Physical Model
 - The Domain Model
 - The Property Graph Model
 - Relational Versus Graph Models
 - Actions
- *Property Graph Databases*
 - Property Graph Model Components
 - Table (FKs)-[:BECOME]->(Relationships)
 - Relationship Properties
 - Graph Query Language: Cypher
 - Representing Data in the Graph
 - Express Complex Relationship Queries



BD-02: Introduction to NoSQL

- Basic Query Results
 - Basic Query Results as a Graph
- *Semantic Graph Databases*
 - Property Graphs vs. Semantic Graphs
 - Semantic Graph
 - RDF Triple Store
 - SPARQL 1.1
 - SPRQL Examples
- *Graph Engines*
 - BSP High Level Overview
 - Bulk Synchronous Parallel (BSP)
 - Graph Engine Dataflow
 - Graph Engine Process Steps
 - Convert Tables to Java Objects in Memory

Module 6. The Future of NoSQL (8 min)

- *Overview*
- *Questions for your NoSQL Prospect Vendor*
- *Future of NoSQL*
- *Big Data and NoSQL Sales Projection*
- *The NoSQL Challenge*
- *Getting Started*
- *What Technology to Select*