Designing Software Architecture for Big Data Systems

Training Program Outline

Duration: 10 Days | 4 Hours/Day | Total: 40 Hours

Day 1: Foundations of Software and Big Data Architecture

- Introduction to Software Architecture in the Big Data Context
- Key Architecture Qualities: Scalability, Modularity, Reliability, Maintainability
- Evolution of Big Data Systems: From Monoliths to Distributed Architectures
- Role of a Business Analyst in Architecture Planning
- Case Study: Basic Big Data System Walkthrough

Day 2: Core Components and Layers in Big Data Architecture

- Data Ingestion, Storage, Processing, and Consumption Layers
- Batch vs. Real-time Architectures
- Architecture Blueprints: Lambda, Kappa, and Delta Architecture
- Aligning Architecture with Business Requirements
- Hands-On: Mapping Business Needs to Architecture Layers

Day 3: Architectural Design Methodologies and Patterns

- Microservices, Service-Oriented Architecture (SOA), and Event-Driven Design
- Choosing the Right Pattern for the Use Case
- Trade-off Analysis: Complexity vs. Maintainability, Flexibility vs. Performance
- Hands-On Workshop: Evaluating Trade-offs in Real Case Scenarios
- Use Case: Telecom Event Stream Processing

Day 4: Technology Stack and Tools in Big Data Systems

- Overview of Modern Big Data Stack (Hadoop, Spark, Kafka, Flink, etc.)
- NoSQL Databases: HBase, Cassandra, MongoDB
- Cloud-Native Technologies: AWS, Azure, GCP (EMR, BigQuery, Databricks)
- Tool Selection Criteria Based on Business Use Cases
- Demo: End-to-End Stack Comparison for a Retail Scenario

Day 5: Designing for Performance and Scalability

- System Bottlenecks and Performance Tuning Approaches
- Caching Strategies (Redis, Memcached), Load Balancing Techniques
- Auto-Scaling and Elastic Compute on Cloud
- Hands-On Lab: Diagnosing Performance Issues in Architecture
- Real-World Example: Scalable System in a Banking Analytics Use Case

Day 6: Containerization, CI/CD, and DevOps Integration

- Containers (Docker), Orchestration (Kubernetes), and Microservices Deployment
- CI/CD Pipelines for Big Data Systems
- Infrastructure as Code (Terraform, CloudFormation)
- Integrating DevOps into Big Data Architectures
- Case Study: Cloud-native Big Data Deployment in Telecom

Day 7: Data Governance, Security, and Compliance

- Data Privacy (GDPR, CCPA), Auditability, and Lineage
- Role-Based Access, Encryption, and Secure APIs
- Metadata Management and Cataloging
- Business Risk Mitigation via Data Governance
- Workshop: Embedding Governance into Architecture Design

Day 8: Mapping Business Requirements to Architecture

- Requirement Gathering Techniques for Big Data Projects
- Translating Real-time Analytics and Predictive Modeling into Architecture
- Business KPIs and their Architectural Implications
- Interactive Session: Creating Architecture Specs from Business Goals
- Practical Exercise: Use Case Modeling (Retail or Healthcare)

Day 9: Real-World Big Data Architecture Reviews

- Deep Dive into Existing Architectures from Retail, Telecom, and Finance
- Identifying Best Practices and Anti-patterns
- Optimization Case Studies
- Group Activity: Reverse Engineering Architecture Diagrams
- Expert Guest Talk (optional): Architecting Big Data at Scale

Day 10: Business Value, ROI, and Strategic Alignment

- Cost Drivers: Compute, Storage, Data Transfer, Support
- ROI Modeling for Architecture Investments
- Architecture's Impact on Business KPIs (Latency, Insight Generation, Uptime)
- Capstone Project: Design and Present a Big Data Architecture Solution
- Final Q&A, Feedback, and Certification/Completion