Streaming Data: Concepts, Applications, and Technologies

by Dave Wells & Kevin Petrie
Module 0. About the Course (5 min)

Module 1. The Need for Speed (21 min)
- Batch Vs. Real Time
- The Speed of Business
- The Speed of Data Parts 1-3
- Fast Data Drives the Consumer World
- Fast Data Drives the Business World
- Fast Data Drives the Analytic World
- The Business Case for Fast Data
  - Daily Operations
  - Loss Prevention Part 1-2
  - Financial Growth
  - Efficiency Gains
  - Opportunity Innovation
- Summary

Module 2. Moving Data through Pipelines (61 min)
- Data Pipeline Processing
  - Data Pipelines in Context
  - Data Pipeline Complexity
  - Data Pipeline Processing Patterns
- Data Stream Processing
  - Use Case Patterns with Stream Processing
  - Stream Processing Example
  - Transportation and Logistics
  - Telecom
  - Oil and Gas
  - Smart Transportation
  - Smart Grid
  - Industrial Automation
  - Internet Security
- IoT and Edge Computing
  - The Internet of Things (IoT)
  - IoT Machines and Sensors
  - IoT Applications Parts 1-3
  - What is Edge Computing? Part 1-2
  - Edge vs Cloud
  - About Microservices
  - Modular Systems vs Microservices
  - Microservices Example
  - IoT and Edge Computing
  - Self-Driving Vehicle Example
  - Revisiting the Gas Pipeline

Module 3. Change Data Capture (10 min)
- Data Management – A New Streaming Use Case
- What is CDC?
- Date Stamps for Change Detection
• Diff Comparisons for Change Detection
• Database Triggers for Change Detection
• In-Database Change Data Capture
• Log Based Change Data Capture
• Log Based CDC for Data Warehousing and Data Lakes

Module 4. Streaming First Architecture for Data Management (17 min)
• Modern Enterprise Data Requirements
• Essential Characteristics of Streaming Technology
• Stream Processing Technologies
• Kafka Stream Processing Basics
• Replacing ETL with CDC and Streaming
• Streaming First Architecture
• Stream to Data Warehousing
• Stream to Slowly Changing Dimensions
• Stream to Data Lake
• Stream to Analytic Applications
• Moving to Streaming First Architecture

Module 5. Data Streaming with Apache Kafka (11 min)
• Introduction to Apache Kafka
  o Apache Kafka – Overview
  o Apache Kafka – Principles
  o Apache Kafka – Inception
• Kafka Predecessors
  o What Came Before
  o Earlier Messaging Systems Part 1-2
  o Limitations of Earlier Messaging Systems
  o Apache Kafka – Key Differences with Predecessors

Module 6. Kafka Architecture (19 min)
• Apache Kafka Architecture
• Essential Components Part 1-2
• Data and Process Flow Part 1-2
• Kafka Record – Key Components
• Kafka Broker Overview
  o Record Writing Process
  o Record Reading Process
• Cluster Management with Apache
  o Kafka Connect
  o Kafka Streams API

Module 7. Data Stream Ecosystems and Use Cases (40 min)
• Data Streaming Use Cases
  o Use Cases Part 1-3
• Streaming Integration with Data Ecosystems
  o Microservices and Kafka
Example of Microservices Enablement
Platform Evolution and Convergence
Examples of Converged Architectures 1-3
Lambda Architecture
Kappa Architectures

- Kafka Alternatives from Cloud Service Providers
- Amazon Kinesis
  - Amazon Kinesis – Overview
  - Amazon Kinesis – Essential Components
  - Amazon Kinesis – Offerings
  - Amazon Kinesis – Data Streaming Use Cases
  - Amazon Kinesis – Additional Use Cases
  - Amazon Kinesis – Example Pipeline
- Azure Event Hub
  - Azure Event Hub – Overview
  - Azure Event Hub – Essential Components
  - Azure Event Hub – Use Cases
- Google Cloud PUB/SUB
  - Google Cloud Pub/Sub – Overview
  - Google Cloud Pub/Sub – Essential Components
  - Google Cloud Pub/Sub – Architecture