

## Kafka Development (32 Hrs)

- **Module I: Introduction (Day 1)**

- Traditional Messaging System Models & Problems. Enterprise Message Bus Concept
- Distributed / Parallel Processing Concepts. Big Data & Streaming Data Characteristics
- Introduction & Architecture Overview, Theory & Fundamentals
- Solution Types, Distributions & Specialties, Challenges & Complexity & Use Cases.

- **Module II: Core Concepts (Day 1)**

- Topics, Partitions and Offsets, Brokers and Topics, Topic Replication
- Producers and Message Keys
- Consumers & Consumer Groups, Consumer Offsets & Delivery Semantics
- Kafka Broker Discovery, Zookeeper, Kafka Guarantees, Message Ordering

- **Module III: Setup (Day 1)**

- Multi Node Confluent Cluster Setup with Zookeeper
- Eclipse Maven Project Setup
- SSH & SCP Tools Overview
- Production Environment High Level Architecture Guidelines & Environment Discussion

- **Module IV: CLI & Core Functionality Hands-on (Day 2)**

- Topics, Console Producer & Consumer, Consumer Groups, Resetting Offsets & Important CLI Options
- Java Producer, Producer Call-backs, Producer with Keys, Synchronous, Asynchronous, Fire & Forget
- Java Consumer, Consumer inside Consumer Group, Multi-threaded Consumer, Consumer Seek & Assign
- Choosing Partition Count & Replication Factor
- Producer Configurations, acks & min.insync.replicas, retries & max.in.flight.requests.per.connection, Idempotence, Type Safety, Compression, Batching, High Throughput, Default Partitions and Key Hashing, max.block.ms, buffer.memory, custom partitioning, multi event model
- Delivery Semantics for Consumers, Consumer Idempotence, Offset Commit Strategies, Manual Commit of Offsets, Performance Improvement using Batching, Offsets Reset Behaviour, Replaying Data, Internal Threads
- Changing Topic Configuration, Segment and Indexes, Log Clean-up & Compaction

- **Module V: Schema Registry (Day 3)**

- Avro overview, Schemas
- Flexible Schemas with JSON / Binary Serialization and defensive programming
- Topic Schema management, Validation of schema
- Prevent Consumer from accepting unexpected schema / defensive programming
- Prevent producers from sending messages that don't align with schema registry

- **Module VI: REST API (Day 3)**

- Using the REST API to write a Producer
- Using the REST API to write a Consumer

- **Module VII: Connect API (Day 3)**

- Concepts, Connect Source & Sink Architecture Design, Connectors, Configuration, Tasks, Workers
- File Stream Source, Sink Connector
- JDBC Source Connector, HDFS Sink Connector

- **Module VIII: Streams (Day 4)**

- Core Concepts, Topologies, Sink, Sources, Processors, Streams
- Kafka Streams vs other stream processing libraries ( Spark Streaming, NiFi, Flink )
- Data processing & Real-Time Data Ingestion (Spark integration - BigData)
- KStreams and KTables Stateful / Stateless Operations
- Map, Filter, Read/Write & Transformations, Exactly Once Semantics
- Window Operations
- Stream Joins

- **Module IX: KSQL (Day 4)**

- Introduction
- Using KSQL for Data Manipulation
- Streams & Tables
- KSQL - Aggregations

- **Module X: Wrap-up (Day 4)**

- Discussing Current Market Standards & New / Incubating Tech. Other Q & A, Doubt Clearance.
- Lab Participation by Candidates on Multiple Domains Simulated Data Patterns for Real Life Project Scenarios.