COURSE TABLE OF CONTENTS

Apache NiFi and Kafka Integration

OEM: Apache Software Foundation • Duration: 5 Days (40 hrs)

COURSE MODULES & TOPICS

Module 1: Introduction to Data Engineering and Apache NiFi

- Data engineering and integration overview
- Batch vs real-time data processing
- Data pipelines and ETL/ELT concepts
- Enterprise data integration challenges
- Apache NiFi features and introduction
- NiFi use cases and industry applications
- Comparison with other data integration tools

Module 2: Apache NiFi Architecture and Installation

- NiFi architecture and design principles
- Core NiFi components
- Flow Controller overview
- Processors, FlowFiles, connections, relationships
- Process groups, input and output ports
- NiFi installation and environment configuration
- Java runtime and NiFi configuration
- Managing NiFi services
- NiFi web-based user interface

Module 3: NiFi Data Flow Development

- FlowFiles and attributes
- Creating and managing process groups
- Adding and configuring processors
- Connecting processors and managing relationships
- Processor scheduling and execution strategies
- Managing queues and prioritization
- Data provenance and FlowFile tracking
- Importing and exporting data flows

Module 4: NiFi Hands-on Practical Exercises

- Creating the first data flow
- Reading files from local directories
- Processing and transferring files
- Configuring processor properties
- Monitoring FlowFiles and queue statistics
- Building a basic end-to-end data flow

Module 5: Working with Core NiFi Processors

- Overview and classification of processors
- File-based processors: GetFile, ListFile, FetchFile, PutFile

© Koenig Solutions

- Database and record-based processors: ExecuteSQL, QueryDatabaseTable, PutDatabaseRecord, ConvertRecord
- HTTP and API integration processors: InvokeHTTP, HandleHttpRequest, HandleHttpResponse
- Processor configuration and property management
- Scheduling and execution configuration

Module 6: Data Transformation and Enrichment in NiFi

- Introduction to data transformation
- Working with FlowFile attributes
- UpdateAttribute, ReplaceText, ExtractText processors
- EvaluateJsonPath and JoltTransformJSON processors
- Transforming and enriching incoming data
- Dynamic attribute creation and expression language

Module 7: Data Routing, Filtering and Validation

- Data routing and filtering introduction
- RouteOnAttribute and RouteOnContent processors
- QueryRecord and ValidateRecord processors
- Conditional routing based on business rules
- Data validation and quality checks
- Handling invalid and rejected records

Module 8: Data Formats and Record Processing

- Structured and semi-structured data overview
- Working with JSON, XML, CSV, Avro formats
- Record readers and writers
- CSVReader, JsonTreeReader, AvroReader, JsonRecordSetWriter
- Schema management and validation
- Data format conversion and normalization
- Hands-on: building transformation and routing pipelines

Module 9: Introduction to Apache Kafka

- Distributed event streaming introduction
- Apache Kafka features overview
- Kafka use cases in enterprise applications
- Kafka vs traditional message queues
- Kafka vs batch-based processing
- Event-driven architecture fundamentals
- Real-time streaming data architecture

Module 10: Apache Kafka Architecture and Components

- Kafka cluster architecture
- Brokers and cluster management
- Topics and partitions
- Replication factor and partition distribution
- Leader and follower replicas
- Controllers and metadata management
- Storage architecture and log segments
- Retention policies and log compaction
- KRaft and ZooKeeper-based deployments

Module 11: Kafka Producers and Consumers

- Kafka producers introduction
- Producer configuration and properties
- Publishing messages to topics
- Producer acknowledgement mechanisms
- Message keys and partitioning strategies
- Kafka consumers introduction
- Consumer groups and coordination
- Offset management and commit strategies
- Message consumption and processing
- Handling consumer rebalancing

Module 12: Kafka Installation and Practical Implementation

- Kafka installation and environment setup
- Configuring Kafka brokers
- Managing Kafka services
- Creating, listing, deleting topics
- Producing messages with console producer
- Consuming messages with console consumer
- Inspecting topic partitions and offsets
- Hands-on: producer-consumer messaging applications

Module 13: NiFi and Kafka Integration Architecture

- NiFi-Kafka integration overview
- Real-time data streaming architecture
- Role of NiFi in Kafka pipelines
- Kafka as a messaging platform
- Data producers, consumers, brokers
- Designing end-to-end integration workflows
- Integration patterns and use cases
- Data flow lifecycle from source to destination

Module 14: Kafka Integration Processors in NiFi

- Overview of Kafka processors in NiFi
- PublishKafka and ConsumeKafka processors
- Kafka broker connection properties
- Configuring topic names and partitions
- Producer and consumer properties
- Configuring consumer group IDs
- Managing acknowledgements and offsets
- Message delivery and consumption strategies

Module 15: Publishing Data from NiFi to Kafka

- Designing ingestion pipelines in NiFi
- Reading data from files, databases, REST APIs
- Transforming data before publishing
- Configuring PublishKafka processor
- Publishing JSON, CSV, Avro data to Kafka
- Managing message keys and attributes

- Implementing dynamic topic routing
- Validating published messages

Module 16: Consuming Kafka Messages Using NiFi

- Configuring ConsumeKafka processor
- Consuming messages from Kafka topics
- Consumer group behaviour
- Managing offsets and commit strategies
- Processing incoming Kafka messages
- Transforming consumed messages
- Routing messages to files, databases, APIs
- Error handling for consumed messages

Module 17: Hands-on NiFi-Kafka Integration

- Building a file-to-Kafka data pipeline
- Publishing real-time JSON data from NiFi to Kafka
- Consuming Kafka messages into NiFi
- Transforming and routing consumed data