TechyEdz Solutions

Training | Consulting | Developement | Outsourcing

Apache Kafka + Cassandra

9032803832

9032803832

contact@techyedz.com

www.techyedz.com

APACHE KAFKA + CASSANDRA COMBO COURSE

COURSE OVERVIEW:

Apache Kafka is an open-source stream processing platform and a high-performance real-time messaging system that can process millions of messages per second. It provides a distributed and partitioned messaging system that is highly fault tolerant. This Kafka Training course will guide participants through Kafka architecture, installation, interfaces and configuration on their way to learning the advanced concepts of Big Data.

Apache Cassandra is a free, open-source project and a second-generation distributed No SQL database and is considered to be the best choice for high availability and scalability databases, particularly when dealing with large amounts of data. Cassandra supports replication across multiple data centres, while also making the write and read processes highly scalable by offering tenable consistency. This Apache Cassandra training course will provide you with an overview of the fundamentals of Big Data and No SQL databases, an understanding of Cassandra and its features, architecture and data model.

COURSE OUTLINE:

APACHE KAFKA

1. GETTING STARTED

- 1.1 Introduction
- 1.2 Use Cases
- 1.3 Quick Start
- 1.4 Ecosystem
- 1.5 Upgrading

2. APIS

- 2.1 Producer API
- 2.2 Consumer API
- 2.3 Streams API
- 2.4 Connect API
- 2.5 Admin API

3. CONFIGURATION

- 3.1 Broker Configs
- 3.2 Topic Configs
- 3.3 Producer Configs
- 3.4 Consumer Configs
- 3.5 Kafka Connect Configs
- 3.6 Kafka Streams Configs
- 3.7 AdminClient Configs

4. DESIGN

- 4.1 Motivation
- 4.2 Persistence
- 4.3 Efficiency
- 4.4 The Producer
- 4.5 The Consumer
- 4.6 Message Delivery Semantics
- 4.7 Replication
- 4.8 Log Compaction
- 4.9 Quotas

5. IMPLEMENTATION

- 5.1 Network Layer
- 5.2 Messages
- 5.3 Message format
- 5.4 Log
- 5.5 Distribution

6. OPERATIONS

- 6.1 Basic Kafka Operations
  - Adding and removing topics
  - Modifying topics
  - Graceful shutdown
  - Balancing leadership
  - Checking consumer position
  - Mirroring data between clusters
  - Expanding your cluster
  - Decommissioning brokers
  - Increasing replication factor
- 6.2 Datacenters

- 6.3 Important Configs
  - Important Client Configs
  - A Production Server Configs
- 6.4 Java Version
- 6.5 Hardware and OS
  - OS
  - Disks and Filesystems
  - Application vs OS Flush Management
  - Linux Flush Behavior
  - Ext4 Notes
- 6.6 Monitoring
  - Selector Monitoring
  - Common Node Monitoring
  - Producer Monitoring
  - Consumer Monitoring
  - Connect Monitoring
  - Streams Monitoring
  - Others
- 6.7 ZooKeeper
  - Stable Version
  - Operationalization

7. SECURITY

- 7.1 Security Overview
- 7.2 Encryption and Authentication using SSL
- 7.3 Authentication using SASL
- 7.4 Authorization and ACLs
- 7.5 Incorporating Security Features in a Running Cluster
- 7.6 ZooKeeper Authentication
  - New Clusters
  - ZooKeeper SASL Authentication
  - ZooKeeper Mutual TLS Authentication
  - Migrating Clusters
  - Migrating the ZooKeeper Ensemble
  - ZooKeeper Quorum Mutual TLS Authentication
- 7.7 ZooKeeper Encryption

8. KAFKA CONNECT

- 8.1 Overview
- 8.2 User Guide
  - Running Kafka Connect
  - Configuring Connectors
  - Transformations
  - REST API
- 8.3 Connector Development Guide

9. KAFKA STREAMS

- 9.1 Play with a Streams Application
- 9.2 Write your own Streams Applications
- 9.3 Developer Manual
- 9.4 Core Concepts
- 9.5 Architecture
- 9.6 Upgrade Guide

Apache Cassandra

Course Outline:

1. Getting Started with the Architecture

- Understanding that Cassandra is a Distributed Database
- View Cassandra documentation
- Learning What Snitch is For
- Learning What Gossip is For
- Check your understanding
- Learning How Data Gets Distributed and Replicated
- Access related documentation
- View the course projects

2. Installing Cassandra

- Downloading Cassandra
- Locate the tarball
- Ensuring Oracle Java is Installed
- Check that Oracle Java is installed
- Installing Cassandra

- Viewing the Main Configuration File
- Providing Cassandra with Permission to Directories
- Starting Cassandra
- Checking Status
- Accessing the Cassandra system.log File

3. Communicating with Cassandra

- Understanding Ways to Communicate with Cassandra
- View CQL commands
- Using cqlsh
- Access cqlsh
- Use cqlsh

4. Creating a Database

- Understanding a Cassandra Database
- View the existing keyspaces
- Defining a Keyspace
- Define a keyspace
- Create a second database

5. Creating a Table

- Creating a Table
- Enter a keyspace
- Defining Columns and Data Types
- Defining a Primary Key
- Recognizing a Partition Key
- Specifying a Descending Clustering Order
- Create a second table

6. Inserting Data

- Understanding Ways to Write Data
- Using the INSERT INTO command
- Using the COPY command
- Seeing How Data is Stored in Cassandra
- Seeing How Data is Stored on Disk
- See how data is stored on disk
- Insert data

7. Modeling Data

- Understanding Data Modeling in Cassandra
- Understanding Secondary Indexes
- Use a WHERE clause
- Creating a Secondary Index
- Defining a Composite Partition Key

8. Creating an Application

- Understanding Cassandra Drivers
- View the Cassandra drivers
- Setting Up a Development Environment
- Access a development environment
- Creating an Application Page
- Executing a Query
- Displaying Query Results
- Lab: Create A Second Application - Part 1
- Lab: Create A Second Application - Part 2
- Lab: Create A Second Application - Part 3

9. Updating and Deleting Data

- Updating Data
- Understanding How Updating Works
- Get an inside view into how updating works
- Deleting Data
- Update and delete data

10. Selecting Hardware

- Understanding Hardware Choices
- Understanding RAM and CPU
- Recommendations
- Selecting Storage
- Deploying In the Cloud

11. Adding Nodes to a Cluster

- Understanding Cassandra Nodes
- Having a Network Connection - Part 1
- Having a Network Connection - Part 2
- Having a Network Connection - Part 3

- Specifying the IP Address Of A Node In
- Cassandra
- Specifying Seed Nodes
- Bootstrapping a Node
- Cleaning Up a Node
- Using Cassandra-stress
- Lab: Add a Third Node

12. Monitoring A Cluster

- Understanding Cassandra Monitoring Tools
- Using Node tool
- Using J Console
- Learning About Ops Centre

13. Repairing Nodes

- Understanding Repair
- Repairing Nodes
- Understanding Consistency - Part 1
- Understanding Consistency - Part 2
- Understanding Hinted Handoff
- Understanding Read Repair
- Lab: Repair Nodes for a Key space

14. Removing a Node

- Understanding Removing a Node
- Decommissioning a Node
- Putting a Node Back Into Service
- Removing a Dead Node
- Lab: Put a Node Back Into Service

15. Redefining a Cluster for Multiple Data Centres

- Redefining For Multiple Data Centres - Part 1
- Redefining For Multiple Data Centres - Part 2
- Changing Snitch Type
- Modifying cassandra-rackdc.properties
- Changing Replication Strategy - Part 1
- Changing Replication Strategy - Part 2

WHO CAN ATTEND

The Apache Kafka course is suitable for many roles, including:

- Professionals aspiring for a career in Big Data
- Analytics professionals, research professionals, IT developers, testers and project managers
- Students
- Individuals looking for a change in career

PREREQUISITES

- Knowledge of any messaging system
- Basic knowledge of Java or any programming language
- Some knowledge of Linux- or Unix-based systems (nor required)

Number of Hours: 75hrs

Key Features:

- One to One Training
- Online Training
- Fastrack & Normal Track
- Resume Modification
- Mock Interviews
- Video Tutorials
- Materials
- Real Time Projects
- Virtual Live Experience
- Preparing for Certification