Real-Time Data Engineering in the Cloud

Course Summary

Description

To handle real-time big data, you need to solve two difficult problems: how do you ingest that much data and how will you process that much data? This course explores the latest real-time frameworks—both open source and managed cloud services—discusses the leading cloud providers, and explains how to choose the right one for your company.

Real-time big data frameworks are enabling brand-new use cases, while the cloud is letting us do things cheaper and faster than ever. Together, they're making it easier to create production real-time systems. But to handle real-time big data, you need to solve two difficult problems: how do you ingest that much data and how will you process that much data? This course explores the latest real-time frameworks—both open source and managed cloud services—discusses the leading cloud providers, and explains how to choose the right one for your company. Focusing on Apache Kafka and Apache Spark, the course also demonstrates how to ingest data, process it, analyze it, and display it in real time with a dashboard.

Objectives
By the end of this course, students will be able to:
- Explore the latest real-time frameworks and the leading cloud providers
- Discover how to choose the right open source technology or managed cloud service for your company
- Learn how to create real-time data pipelines using the latest open source and managed cloud services, ingest data in real time and at scale, process the data once you've ingested it, and create data products from real-time data sources

Topics

- Real-time Data Engineering
- Real-time Data Pipelines
- Real-time Technologies
- Real-time Pipelines
- Pros and Cons of Real-time
- Using the Cloud
- Cloud Providers
- Real-time Technologies
- Choosing a Provider
- Ingesting Data
- Real-time Ingestion
- Real-time ETL5
- Kafka
- About Kafka

- Kafka Internals
- Kafka API
- Processing Data
- Real-time Data Processing
- Real-time Processing Technologies
- Spark Streaming
- Spark Streaming
- Streaming API
- Advanced Streaming
- Data Products
- Analysis of Data
- Dashboarding
- Conclusion

Audience

This course is aimed at Data Engineers, Software Engineers, Big Data Engineers, DevOps Engineers, and Software Architects.

Prerequisites

This course is designed for students with strong Java skills.

Duration

Two days