Building Real Time Streaming Systems With Kafka + Spark + Cassandra

Course Summary

Description

This course will teach students on how to build streaming systems using the popular fast data stack: Apache Kafka + Apache Spark + Apache Cassandra.

Objective

- Kafka (1 day)
- Cassandra (1.5 days)
- Spark (1.5 days)
- Putting it all together (1 day)

Topics

- Kafka (1 day)
- Cassandra (1.5 days)
- Spark (1.5 days)
- Putting it all together (1 day)
- Bring Your Own Use Case – Group Study (time permitting)

Audience

This course is designed for Developers / Architects.

Prerequisite

- Familiarity with either Java / Scala language (our labs in Scala and Java – we provide a quick Scala introduction)
- Basic understanding of Linux development environment (command line navigation / running commands)
- No previous knowledge of Kafka / Spark / Cassandra is assumed. The course will cover all the technologies and teach them how to integrate them.

Duration

Five Days
Building Real Time Streaming Systems With Kafka + Spark + Cassandra

Course Outline

I. **Kafka (1 day)**
   A. Kafka design & architecture
   B. Getting Kafka up and running
   C. Using Kafka utilities
   D. Reading & Writing to Kafka using Java API
      - Labs: all of the above sections

II. **Cassandra (1.5 days)**
   A. Cassandra design & architecture
   B. CQLSH
   C. Read / Write path in Cassandra
   D. C* eventual consistency
   E. Time series data
   F. Data modeling on C*
   G. Using C* Java API
      - Labs: all of the above sections

III. **Spark (1.5 days)**
   A. Scala primer (if required)
   B. Spark design and architecture
   C. Spark Shell
   D. Spark Data structures: RDDs, Dataframes, Datasets
   E. Batch analytics with Spark
   F. Writing Spark applications using Spark APIs
   G. Spark streaming
   H. Structured streaming
      - Labs: all of the above sections

IV. **Putting it all together (1 day)**
   A. Reading Kafka streams from Spark
   B. Saving streaming data from Spark into Cassandra
   C. Full end to end application
   D. Benchmarking
   E. Monitoring
   F. Tuning and Optimizing the system
      - Labs: all of the above sections

V. **Bring Your Own Use Case – Group Study (time permitting)**
   - We encourage students to bring a use case they are working on at their company for discussion with the class
   - We will discuss the use case in the class
   - Discuss design choices, sketch out a few designs, debate pros/cons of each design
   - Discuss best practices
   - This will be a group activity, and will be a lots of fun!