**Introduction to Kafka**

**Overview**

This course will teach Apache Kafka – a popular distributed messaging system– to students.

This course is intended for developers.

This is an **introductory – intermediate course**.

**No previous knowledge of Kafka is expected**.

This class is taught using Java language.

**What You Will Learn**

* Overview of Streaming technologies
* Kafka concepts and architecture
* Programming using Kafka API
* Kafka Connect
* Kafka Streams
* Monitoring Kafka
* Tuning / Troubleshooting Kafka

**Audience :**

Developers

**Skill Level:**Introductory - Intermediate

**Duration :**

two days

**Format :**

Lectures and hands on labs. (50%   50%)

**Prerequisites**

* comfortable with Java programming language (programming exercises are in java)
* comfortable in Linux environment (be able to navigate Linux command line, run commands)

**Lab environment**

**Zero Install :** There is no need to install hadoop software on students’ machines! A lab environment in the cloud will be provided for students.

Students will need the following

* a SSH client (Linux and Mac already have ssh clients, for Windows [Putty](http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html) is recommended)
* a browser to access the cluster. We recommend Chrome browser

**Detailed outline**

* **Section 1: Introduction to Streaming Systems**
  + Fast data
  + Streaming architecture
  + Lambda architecture
  + Message queues
  + Streaming processors
* **Section 2: Introduction**
  + Architecture
  + Comparing Kafka with other queue systems (JMS / MQ)
  + Kaka concepts : Messages, Topics, Partitions, Brokers, Producers, commit logs
  + Kafka & Zookeeper
  + Producing messages
  + Consuming messages (Consumers, Consumer Groups)
  + Message retention
  + Scaling Kafka
  + Labs :
    - Getting Kafka up and running
    - Using Kafka utilities
* **Section 3 : Programming With Kafka**
  + Configuration parameters
  + Producer API (Sending messages to Kafka)
  + Consumer API (consuming messages from Kafka)
  + Commits , Offsets, Seeking
  + Kafka Connect:
    - Connect eco system for HDFS, Elastic Search
    - Sample configurations
  + Schema with Avro
  + Lab :
    - Writing Kafka clients in Java
    - Benchmarking Producer APIs
* **Section 4: Kafka Streams**
  + Streams overview and architecture
  + Streams use cases and comparison with other platforms
  + Learning Kafka Streaming concepts (KStream, KTable, KStore)
  + KStreaming operations (transformations, filters, joins, aggregations)
  + Labs:
    - Kafka Streaming labs
* **Section 5 : Administering Kafka (Quick overview)**
  + Hardware / Software requirements
  + Deploying Kafka
  + Configuration of brokers / topics / partitions / producers / consumers
  + Security: How secure Kafka cluster, and secure client communications (SASL, Kerberos)
  + Monitoring : monitoring tools
  + Capacity Planning : estimating usage and demand
  + Trouble shooting : failure scenarios and recovery
* **Section 6: Monitoring and Instrumenting Kafka**
  + Monitoring Kafka
  + Instrumenting with Metrics library
  + Labs
    - Monitor Kafka cluster
    - Instrument Kafka applications and monitor their performance
* **Final workshop (time permitting)**
  + students will build an end-to-end application simulating web traffic and send metrics to Grafana.  
    See screenshot and video below!

