

Advanced Workshop on IoT/Big-Data Analytic

Summary

The Internet of Things (IoT) and the real time data analytics promises many new technological innovations and business benefits. The success of any solution leveraging these technologies lies in the ability to process and analyze the vast amounts of data produced by the millions or even billions of embedded devices, sensors, appliances and other data-collecting systems in real time. It requires new processes and tools for collecting, storing and processing IoT big data. This workshop introduces the data and analytic flows with a specific focus on IoT and real-time processing of event/streaming data.

The workshop first defines IoT and why IoT data processing is very different from typical data analytics, with its unique requirements for big data and real-time processing. Using a hands-on approach with simulated data, participants will learn to build a messaging and data streaming system with Apache Spark and Kafka and perform a real-time analysis with IoT and streaming data.

By the end of the workshop, participants will have learned:

1. The characteristics and requirements of IoT specific and streaming data
2. How to build a data flow to connect an IoT system or device data to a Big Data platform in specific formats
3. How to use Big data tools to process IoT and streaming data in distributed computing
4. How to use machine learning algorithms to analyze data and extract intelligence.

Skills Needed

Software installation and some programming experience in R, Java or Python (one of the three) is required.

Tentative Agenda

Day	Time	Topic
1	08:30 – 09:00	Registration
	09:00 – 10:30	Introduction to Big Data & IoT Analytics Problem Scope. Analysis of Large Scale Real-Time and Streaming Data
	10:30 – 11:00	<i>Coffee Break</i>
	11:00 – 13:00	Introduction to Kafka
	13:00 – 14:00	<i>Lunch</i>
	14:00 – 16:00	Lab – Install & Verify Docker Environment for Kafka
	16:00 – 16:30	<i>Coffee Break</i>
	16:30 – 18:30	Lab – Creating Topics & Passing Messages. Group Discussion on Problem
2	09:00 – 10:30	Design of Kafka topics and partitions
	10:30 – 11:00	Coffee Break

	11:00 – 12:30	Lab – Designing topics and partitions
	12:30 – 13:00	Evaluation of the designs and suggested solutions
	13:00 – 14:00	Lunch
	14:00 – 16:00	Lab - Implement Topics and Partitions
	16:00 – 16:30	Coffee Break
	16:30 – 18:30	Lab - Implement Topics and Partitions
3	09:00 – 10:30	Introduction to Spark / Spark Streaming
	10:30 – 11:00	Coffee Break
	11:00 – 13:00	Real-time data processing using Kafka and Spark streaming
	13:00 – 14:00	Lunch
	14:00 – 16:00	Lab - Real-time data processing using Kafka and Spark Streaming
	16:00 – 16:15	Coffee Break
	16:15 – 18:00	Lab - Real-time data processing using Kafka and Spark Streaming
4	09:00 – 10:30	Introduction to IoT & Analytics Processing
	10:30 – 11:00	Coffee Break
	11:00 – 13:00	Lab – IoT Kafka Streaming Solution
	13:00 – 14:00	Lunch
	14:00 – 16:00	Lab – IoT Kafka Streaming Solution
	16:00 – 16:30	Coffee Break
	16:30 – 18:30	Lab – IoT data Analysis: Project Discussion
5	09:00 – 10:00	Real-time Sentiment Data Analysis
	10:00 – 10:30	Lab – Twitter Stream Sentiment Analysis
	10:30 – 11:00	Coffee Break
	11:00 – 13:00	Lab - Twitter Stream Sentiment Analysis
	13:00 – 14:00	Lunch
	14:00 – 16:00	Lab – Twitter Stream Sentiment Analysis
	16:00 – 16:30	Coffee Break
	16:30 – 18:00	Recap & Course Close