

TECHNOLOGY TRANSFER PRESENTS

JESSE ANDERSON

REAL-TIME BIG DATA SYSTEMS WITH SPARK STREAMING AND KAFKA

ONLINE LIVE STREAMING

OCTOBER 11-12, 2022



info@technologytransfer.it
www.technologytransfer.it

ABOUT THIS SEMINAR

Takes a participant through the benefits and challenges of real-time Big Data systems. We cover real-time Big Data services that are open source or managed services from Cloud providers. The class focuses on Apache Kafka and Apache Spark Streaming.

It shows how to create consumers and publishers in Kafka. Then, we see how to use Apache Spark Streaming to process the data in Kafka and send it back to Kafka. Finally, the data is visualized in real-time on a webpage using Kafka REST.

BENEFITS OF ATTENDING

- How to create large scale real-time systems using both Apache Kafka and Apache Spark Streaming
- How real-time distributed systems are different from batch systems
- How to create Kafka producers and consumers
- How to process data in Kafka with Spark Streaming and place the results back into Kafka
- How to visualize data and show data in real-time on a Web page

AUDIENCE

- Technical
- Software Engineers
- QA
- Analysts

Prerequisites: Intermediate-Level Java

Please, bring your laptop to the workshop

We will send to the attendees the specification for the setup one week before the seminar date.

Technologies Covered

In-depth Coverage

- Apache Spark Streaming
- Apache Kafka

Covered

- Amazon Web Services
- Microsoft Azure
- Google Cloud
- IBM SoftLayer
- Amazon Kinesis
- Microsoft Event Hubs
- Google Pub/Sub
- Apache NiFi
- Apache Flink

OUTLINE

1. Real-time Data Pipelines

This session shows the general types of real-time technologies. We cover common issues associated with real-time technologies and projects. We consider the pros and cons of real-time distributed systems compared to batch distributed systems.

- Real-time Technologies
- Real-time Pipelines
- Pros and Cons of Real-time

2. Using the Cloud

This session shows what the Cloud is and why companies are moving to the Cloud. Considers the top Cloud Providers and niche Big Data Cloud providers. Covers many of the real-time Big Data offerings from Cloud Providers and open source alternatives.

- Cloud Providers
- Technologie real-time
- Choosing a Provider

3. Ingesting Data

This session shows what real-time ingestion is and what technologies are applicable to real-time ingestion. Discusses the difference between real-time and batch ETL. Covers the first mile and last mile issues when dealing with real-time Big Data.

- Real-time Ingestion
- Real-time ETL

4. Kafka

This session teaches the basics of Kafka and how it works. Covers how to create a simple producer and consumer using the Kafka API.

- About Kafka
- Kafka Internals
- Kafka API

5. Processing Data

This session shows what real-time processing is and what technologies are applicable to real-time processing. Discusses the differences between real-time, event-at-a-time, micro-batching, and batch processing. Covers the tradeoffs of real-time processing systems such as: backpressure, push/pull models, and failovers.

- Real-time Data Processing
- Real-time Processing Technologies

6. Spark Streaming

This session teaches the how to use Spark Streaming. Shows how to create real-time Spark ETL and transform data in real-time.

- Spark Streaming
- Streaming API
- Streaming avanzato

7. Data Products

This session teaches how to create the data products that are consumed by the rest of the organization. Shows how to analyze the data. Then, how to take that analysis and create a real-time dashboard to visualize the data.

- Analysis of Data
- Dashboarding

INFORMATION

<p>PARTICIPATION FEE</p> <p>€ 1100</p> <p>The fee includes all seminar documentation.</p> <p>SEMINAR TIMETABLE</p> <p>9.30 am - 1.00 pm 2.00 pm - 5.00 pm</p>	<p>HOW TO REGISTER</p> <p>You must send the registration form with the receipt of the payment to: info@technologytransfer.it</p> <p>TECHNOLOGY TRANSFER S.r.l. Piazza Cavour, 3 - 00193 Rome (Italy)</p> <p>PAYMENT</p> <p>Wire transfer to: Technology Transfer S.r.l. Banca: Credit Agricole Agenzia 1 di Roma IBAN Code: IT 03 W 06230 03202 000057031348 BIC/SWIFT: CRPPIT2P546</p>	<p>GENERAL CONDITIONS</p> <p>DISCOUNT</p> <p>The participants who will register 30 days before the seminar are entitled to a 5% discount.</p> <p>If a company registers 5 participants to the same seminar, it will pay only for 4.</p> <p>Those who benefit of this discount are not entitled to other discounts for the same seminar.</p> <p>CANCELLATION POLICY</p> <p>A full refund is given for any cancellation received more than 15 days before the seminar starts. Cancellations less than 15 days prior the event are liable for 50% of the fee. Cancellations less than one week prior to the event date will be liable for the full fee.</p> <p>CANCELLATION LIABILITY</p> <p>In the case of cancellation of an event for any reason, Technology Transfer's liability is limited to the return of the registration fee only.</p>
---	---	--

JESSE ANDERSON
REAL-TIME BIG DATA SYSTEMS WITH SPARK STREAMING AND KAFKA

October 11-12, 2022

Registration fee: € 1100

If registered participants are unable to attend, or in case of cancellation of the seminar, the general conditions mentioned before are applicable.

first name

surname

job title

organisation

address

postcode

city

country

telephone

fax

e-mail



Stamp and signature

Send your registration form with the receipt of the payment to:
Technology Transfer S.r.l.
Piazza Cavour, 3 - 00193 Rome (Italy)
Tel. +39-06-6832227 - Fax +39-06-6871102
info@technologytransfer.it
www.technologytransfer.it



SPEAKER

Jesse Anderson is a data engineer, creative engineer, and managing director of the Big Data Institute. Mr. Anderson trains employees on Big Data-including cutting-edge technology like Apache Kafka, Apache Hadoop, and Apache Spark. He has taught thousands of students at companies ranging from startups to Fortune 100 companies the skills to become data engineers. He is widely regarded as an expert in the field and recognized for his novel teaching practices. Mr. Anderson is published by O'Reilly and Pragmatic Programmers and has been covered in such prestigious media outlets as the Wall Street Journal, CNN, BBC, NPR, Engadget, and Wired.