

TECHNOLOGY TRANSFER PRESENTS

VLADIMIR BACVANSKI

**PRACTICAL BIG DATA
WITH SPARK WORKSHOP**

MAY 17-19, 2017
RESIDENZA DI RIPETTA - VIA DI RIPETTA, 231
ROME (ITALY)



info@technologytransfer.it
www.technologytransfer.it

ABOUT THIS SEMINAR

In this 3-day course, you will learn how to tame Big Data with Apache Spark. Spark is the fastest growing Big Data system and provides the solid foundation for processing of large volumes of data. We introduce the key concepts of Spark, its architecture and the development model. We show how to use the several APIs provided by Spark to ingest data, process it in parallel on a cluster and control the various aspects of its execution.

This is a practical, hands-on course with numerous exercises. Expect to spend about 50% of your time developing. If you are a bit rusty when it comes to programming, don't worry: we provide full guidance and worked out solutions.

The background of the course: this course and his predecessors has been used to train Big Data Professionals in leading financial technology organizations worldwide and has been delivered as a tutorial at the top industry conferences.

AUDIENCE

- Big Data Developers
- Big Data Architects
- Data Engineers and other Data Professionals

PREREQUISITES

- Familiarity with data technologies such as SQL is required
- Some development experience is recommended

You will need to bring your laptop. We provide a virtual machine that contains Spark, the tools and the code for labs, solutions, and data for the exercises.

OBJECTIVES

Upon completion, students will be able to:

- Identify key parts of Spark architecture and their roles
- Ingest data and process them in parallel
- Develop batch and streaming applications for Spark
- Work with SQL for Spark
- Understand the key concepts behind Machine Learning and GraphX libraries in Spark

OUTLINE

1. Spark Architecture

- What is Spark?
- History of Spark
- Spark & Clusters
- Spark Shell
- Programming Model
- Deploying Spark applications

2. Scala for Spark: A Crash Course

- Scala as a language for Big Data processing
- Immutability
- Collections
- Higher-Order Functions and data processing
- Classes and Traits
- Tuples
- Options

3. Developing with RDDs

- What is an RDD?
- Bootstrapping RDDs
- Common Operations
- Laziness

4. Developing Spark Applications with Scala APIs

- Set Operations
- Caching RDDs
- Using Case Classes with RDDs
- Other RDDs
- Pair RDDs
- Shared Variables

5. Test-Driven Development (TDD) with Spark

- Test-Driven Development with Scala
- Functional Composition
- Externalizing Spark Functions
- TDD in Spark

6. Spark Streaming

- Streaming v. Batch Programs
- Streaming Context
- DStreams
- Stateful Processing
- Window Operations
- Join & Transform Operations
- Output Operations
- Design Patterns for Custom Output Operations
- Caching DStreams
- Checkpointing
- Deploying Spark Streaming Applications

7. Spark SQL and Dataframes

- Contexts
- DataFrame
- Data Sources
- User-Defined Functions
- Integration/Tools

8. Overview of MLib, GraphX, and other Spark APIs

- MLib: Machine Learning Library
- Types
- Statistics
- Other Low-Level Capabilities
- ML Pipelines
- GraphX: Graph Processing
- Property Graph
- Graph Operators
- Graph Algorithms

9. Conclusion

INFORMATION

<p>PARTICIPATION FEE</p> <p>€ 1700</p> <p>The fee includes all seminar documentation, luncheon and coffee breaks.</p> <p>VENUE</p> <p>Residenza di Ripetta Via di Ripetta, 231 Rome (Italy)</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: TECHNOLOGY TRANSFER S.r.l. Piazza Cavour, 3 - 00193 Rome (Italy) Fax +39-06-6871102</p> <p>within May 2, 2017</p> <p>PAYMENT</p> <p>Wire transfer to: Technology Transfer S.r.l. Banca: Cariparma 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. 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>
--	--	---

VLADIMIR BACVANSKI PRACTICAL BIG DATA WITH SPARK WORKSHOP

May 17-19, 2017
Residenza di Ripetta
Via di Ripetta, 231
Rome (Italy)

Registration fee:
€ 1700 (+iva)

If anyone registered is 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

Software has always been **Dr. Vladimir Bacvanski** passion. Developing software faster, better, and scaling it to handle the largest data volumes arriving with great speed is his world. Realizing that great software requires great minds, Dr. Bacvanski also creates teams of exceptional developers and creators who can confidently take on the most challenging software problems. He is founder of SciSpike. His technical interests these days are centered around two main areas: architectures for high performance modular software and Big Data systems. This includes systems built with Node.js, Scala, Akka, commonly deployed in the Cloud. On the Big Data side these are Hadoop and its descendants such as Spark and Flink. He has over twenty years of experience with software technologies in areas such as architecture and design of mission critical and distributed enterprise systems, rule-based systems and languages, modeling tools, real-time systems, agent systems, and database technologies. Dr. Bacvanski also has extensive experience in software architecture and requirements analysis, and has helped many companies to select, transition, and apply new software technologies. He is published worldwide and is a frequent speaker, session chair, and workshop organizer at leading industry events. He is also the recipient of a number of prestigious academic scholarships and grants. Dr. Bacvanski is known for liking to explain how complex stuff works, and is a sought presenter and keynote speaker at leading industry events. O'Reilly recently published his course on the architecture of Big Data and NoSQL systems. For his contributions to the information management community, IBM awarded him the title of the IBM Information Management Champion for the sixth year running on 2015.

Mr. Bacvanski received a Doctoral degree in Computer Science from Aachen University of Technology (RWTH Aachen) in Germany.