Introduction to Apache Spark | Hands-on Spark for Big Data & Machine Learning (TTSK7503)

Category: AI, Data & Analytics | **Vendor:** Trivera Tech

Duration: 24.00 hours (3 days) **19.5 CPD Hours** **Rating:** ⭐ 4.6 (5,878 reviews)

Course Information

Delivery Format: Instructor Led - Online

Course Overview

Introduction to Apache Spark for Big Data & Machine Learning®provides students with a comprehensive, hands-on exploration of enterprise-grade Spark programming, interacting with the significant components mentioned above to craft complete data science solutions.® Students will leave this course armed with the skills they require to begin working with Spark in a practical, real world environment.

About This Course

Apache Spark, a significant component in the Hadoop Ecosystem, is a cluster computing engine used in Big Data. Building on top of the Hadoop YARN and HDFS ecosystem, it offers order-of-magnitude faster processing for many in-memory computing tasks compared to Map/Reduce. It can be programmed in Java, Scala, Python, and R - the favorite languages of Data Scientists - along with SQL-based front ends. With advanced libraries like Mahout and MLib for Machine Learning, GraphX or Neo4J for rich data graph processing as well as access to other NOSQL data stores, Rule engines and other Enterprise components, Spark is a lynchpin in modern Big Data and Data Science computing.

Geared for experienced developers, Introduction to Apache Spark for Big Data & Machine Learning provides students with a comprehensive, hands-on exploration of enterprise-grade Spark programming, interacting with the significant components mentioned above to craft complete data science solutions. Students will leave this course armed with the skills they require to begin working with Spark in a practical, real world environment.

This course is offered in support of the Python programming language but can also be offered for R or Java with advance notice and planning. Our team will work with you to coordinate the languages, tools and environment that will work best for your organization and needs. Please inquire for details.

Who Should Attend

This foundation-level course is geared for intermediate skilled, experienced Developers and Architects (with basic Python experience) who seek to be proficient in advanced, modern development skills working with Apache Spark in an enterprise data environment.

Prerequisites & Entry Requirements

General Prerequisites:

This foundation-level course is geared for intermediate skilled, experienced Developers and Architects (with basic Python experience) who seek to be proficient in advanced, modern development skills working with Apache Spark in an enterprise data environment. TTPS4800 Introduction to Python Programming

TTSQLB3 Introduction to SQL (Basic familiarity is needed, not in-depth SQL skills)

Learning Outcomes

Upon successful completion of this course, participants will be able to:

This ®"skills-centric® course is about 50% hands-on lab and 50% lecture, designed to train attendees in core big data/ Spark development and use skills, coupling the most current, effective techniques with the soundest industry practices. Throughout the course students will be led through a series of progressively advanced topics, where each topic consists of lecture, group discussion, comprehensive hands-on lab exercises, and lab review.This course provides indoctrination in the practical use of the umbrella of technologies that are on the leading edge of data science development focused on Spark and related tools. Working in a hands-on learning environment, students will explore:

-Spark Ecosystem

-Spark Shell

-Spark Data structures (RDD, DataFrame, Dataset)

-Spark SQL

-Modern data formats and Spark

-Spark API

-Spark & Hadoop & Hive

-Spark ML overview

-GraphX

-Time-permitting: Spark Streaming

-Time-permitting: Optional Capstone Workshop (Time-Permitting)

Detailed Course Outline

Spark Introduction

Big data, Hadoop, Spark

Spark concepts and architecture

Spark components overview

Labs: installing and running Spark

The first look at Spark

Spark shell

Spark web UIs

Analyzing dataset €œ part 1

Labs: Spark shell exploration

Spark Data structures

Partitions

Distributed execution

Operations: transformations and actions

Labs: Unstructured data analytics using RDDs

Caching

Caching overview

Various caching mechanisms available in Spark

In memory file systems

Caching use cases and best practices

Labs: Benchmark of caching performance

DataFrames and Datasets

DataFrames Intro

Loading structured data (JSON, CSV) using DataFrames

Using schema

Specifying schema for DataFrames

Labs: DataFrames, Datasets, Schema

Spark SQL

Spark SQL concepts and overview

Defining tables and importing datasets

Querying data using SQL

Handling various storage formats: JSON, Parquet, ORC

Labs: querying structured data using SQL; evaluating data formats

Spark and Hadoop

Hadoop Primer: HDFS, YARN

Hadoop + Spark architecture

Running Spark on Hadoop YARN

Processing HDFS files using Spark

Spark & Hive

Spark API

Overview of Spark APIs in Scala / Python

The lifecycle of a Spark application

Spark APIs

Deploying Spark applications on YARN

Labs: Developing and deploying a Spark application

Spark ML Overview

Machine Learning primer

Machine Learning in Spark: MLib / ML

Spark ML overview (newer Spark2 version)

Algorithms overview: Clustering, Classifications, Recommendations

Labs: Writing ML applications in Spark

GraphX

GraphX library overview

GraphX APIs

Create a Graph and navigating it

Shortest distance

Pregel API

Labs: Processing graph data using Spark

Time Permitting Topics

Spark Streaming

Streaming concepts

Evaluating Streaming platforms

Spark streaming library overview

Streaming operations

Sliding window operations

Structured Streaming

Continuous streaming

Spark & Kafka streaming

Labs: Writing spark streaming applications

Workshop

Attendees will work on solving real-world data analysis problems using Spark

Additional Course Details

Nexus Humans Introduction to Apache Spark | Hands-on Spark for Big Data & Machine Learning (TTSK7503) training program is a workshop that presents an invigorating mix of sessions, lessons, and masterclasses meticulously crafted to propel your learning expedition forward.

This immersive bootcamp-style experience boasts interactive lectures, hands-on labs, and collaborative hackathons, all strategically designed to fortify fundamental concepts.

Guided by seasoned coaches, each session offers priceless insights and practical skills crucial for honing your expertise. Whether you're stepping into the realm of professional skills or a seasoned professional, this comprehensive course ensures you're equipped with the knowledge and prowess necessary for success.

While we feel this is the best course for the Introduction to Apache Spark | Hands-on Spark for Big Data & Machine Learning (TTSK7503) course and one of our Top 10 we encourage you to read the course outline to make sure it is the right content for you.

Additionally, private sessions, closed classes or dedicated events are available both live online and at our training centres in Dublin and London, as well as at your offices anywhere in the UK, Ireland or across EMEA.

Frequently Asked Questions

Q: What delivery options are available for Introduction to Apache Spark | Hands-on Spark for Big Data & Machine Learning (TTSK7503)?

We offer multiple delivery formats:

- Live Instructor-Led Classroom Online (Virtual/Live Online)
- Traditional Instructor-Led Classroom Training (ILT)
- On-site delivery at your offices anywhere in United Kingdom
- Private dedicated courses customized for your team

Q: How many CPD hours does this course provide?

The 3-day Introduction to Apache Spark | Hands-on Spark for Big Data & Machine Learning (TTSK7503) course provides up to 19.5 CPD hours of structured learning. CPD certificates can be provided upon request.

Q: What is the duration of the Introduction to Apache Spark | Hands-on Spark for Big Data & Machine Learning (TTSK7503) training?

The training takes place over 3 day(s), with each day lasting approximately 24.00 hours including breaks for lunch and refreshments.

Q: Do you provide corporate training for Introduction to Apache Spark | Hands-on Spark for Big Data & Machine Learning (TTSK7503)?

Yes, we provide corporate training, dedicated training, and closed classes for Introduction to Apache Spark | Hands-on Spark for Big Data & Machine Learning (TTSK7503). Training can take place anywhere in United Kingdom including London, Manchester, Birmingham, Edinburgh, or live online allowing teams from across United Kingdom or internationally to attend.

Q: What related terms do people search for?

Popular related searches include: Spark

Q: Why choose Nexus Human for Introduction to Apache Spark | Hands-on Spark for Big Data & Machine Learning (TTSK7503)?

Nexus Human is recognized as one of the leading training providers. Our trainers have won multiple awards including:

- Small Firms Best Trainer Award
- National Training Partner of the Year (Ireland) - Multiple Years
- Global Top 30 Instructor Awards (2012, 2019, 2021)
- Tech Excellence Award Nominations
- Learning Performance Institute (LPI) External Training Provider Sponsor 2024

Q: Are there any discount codes available?

Yes! Use discount code **PENPAL5** when booking your Introduction to Apache Spark | Hands-on Spark for Big Data & Machine Learning (TTSK7503) training. Please note that only one discount code can be used per booking and cannot be combined with other special offers.

Nexus Human

Professional Training & Development

Email: info@nexushuman.com

Website: www.nexushuman.com

Phone: +353 1 889 8700 (Ireland) | +44 20 8142 8961 (UK)

© 2026 Nexus Human. All rights reserved. This brochure was generated on 14/11/2026.