Mastering Scala with Apache Spark for the Modern Data Enterprise (TTSK7520)

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

Duration: 40.00 hours (5 days) **32.5 CPD Hours** **Rating:** ⭐ 4.6 (5,878 reviews)

Course Information

Delivery Format: Instructor Led - Online

Course Overview

®Mastering Scala with Apache Spark for the Modern Data Enterprise®is a five day hands-on course designed to provide you with the essential skills and tools to tackle complex data projects using Scala programming language and Apache Spark, a high-performance data processing engine. Mastering these technologies will enable you to perform a wide range of tasks, from data wrangling and analytics to machine learning and artificial intelligence, across various industries and applications.

About This Course

Embark on a journey to master the world of big data with our immersive course on Scala and Spark! Mastering Scala with Apache Spark for the Modern Data Enterprise is a five day hands-on course designed to provide you with the essential skills and tools to tackle complex data projects using Scala programming language and Apache Spark, a high-performance data processing engine. Mastering these technologies will enable you to perform a wide range of tasks, from data wrangling and analytics to machine learning and artificial intelligence, across various industries and applications.

Guided by our expert instructor, you'll explore the fundamentals of Scala programming and Apache Spark while gaining

valuable hands-on experience with Spark programming, RDDs, DataFrames, Spark SQL, and data sources. You'll also explore Spark Streaming, performance optimization techniques, and the integration of popular external libraries, tools, and cloud platforms like AWS, Azure, and GCP. Machine learning enthusiasts will delve into Spark MLlib, covering basics of machine learning algorithms, data preparation, feature extraction, and various techniques such as regression, classification, clustering, and recommendation systems.

You'll also gain experience working with graph processing using Spark GraphX, as well as innovative generative AI

technologies, integrating GPT with Spark and Scala for practical applications. Time permitting, you will also be introduced to Spark NLP, covering text preprocessing, classification, and sentiment analysis. With a focus on practical skills and best practices, you'll work on interesting learning objectives and gain hands-on experience with innovative tools in a live, interactive environment.

Upon completing this course, you'll be ready to confidently apply your newly acquired Scala and Apache Spark skills to a wide range of projects. You'll be able to develop efficient and scalable applications, harness the power of machine learning, and analyze large datasets, giving you a competitive edge in the rapidly evolving world of big data and analytics. By integrating these technologies into your daily work, you'll be better prepared to solve complex problems, streamline processes, and ultimately drive value for your organization.

Who Should Attend

This intermediate and beyond level course is geared for experienced technical professionals in various roles, such as developers, data analysts, data engineers, software engineers, and machine learning engineers.

Practical programming experience is required.

Prerequisites & Entry Requirements

General Prerequisites:

Basic understanding of Java programming: -Familiarity with Java syntax, data structures, and concepts, such as variables,
loops, and conditionals.
- Fundamental knowledge of object-oriented programming (OOP): Experience with OOP principles, such as inheritance,
encapsulation, and polymorphism, in any programming language.
- Familiarity with data structures and algorithms: A basic grasp of common data structures, such as arrays, lists, and maps,
as well as an understanding of simple algorithms, like sorting and searching.
- Experience with distributed systems: Basic awareness of distributed computing concepts, such as data partitioning,
parallel processing, and fault tolerance.
- Basic knowledge of databases: Understanding of database concepts, including data storage, querying, and manipulation
using SQL or NoSQL databases.

Learning Outcomes

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

Working in a hands-on learning environment led by our expert instructor you'll:
- Develop a basic understanding of Scala and Apache Spark fundamentals, enabling you to confidently create scalable and high-performance applications.
- Learn how to process large datasets efficiently, helping you handle complex data challenges and make data-driven
decisions.
- Gain hands-on experience with real-time data streaming, allowing you to manage and analyze data as it flows into your applications.
- Acquire practical knowledge of machine learning algorithms using Spark MLlib, empowering you to create intelligent applications and uncover hidden insights.
- Master graph processing with GraphX, enabling you to analyze and visualize complex relationships in your data.
- Discover generative AI technologies using GPT with Spark and Scala, opening up new possibilities for automating content generation and enhancing data analysis.

Detailed Course Outline

Introduction to Scala

Brief history and motivation

Differences between Scala and Java

Basic Scala syntax and constructs

Scala's functional programming features

Introduction to Apache Spark

Overview and history

Spark components and architecture

Spark ecosystem

Comparing Spark with other big data frameworks

Lab: Practice basic Scala syntax and functional programming concepts using the REPL.

Setting up the Development Environment

Basics of Spark Programming SparkContext and SparkSession

Resilient Distributed Datasets (RDDs)

Transformations and Actions

Working with DataFrames

Spark SQL and Data Sources

Spark SQL library and its advantages

Structured and semi-structured data sources

Reading and writing data in various formats (CSV, JSON, Parquet, Avro, etc.)

Data manipulation using SQL queries

Lab: Setting up the Environment and Running a Simple Spark Application.

Lab: Load and query data from different data sources using Spark SQL.

Data Processing and Spark Programming

Basic RDD Operations

Creating and manipulating RDDs

Common transformations and actions on RDDs

Working with key-value data

Basic DataFrame and Dataset Operations

Creating and manipulating DataFrames and Datasets

Column operations and functions

Filtering, sorting, and aggregating data

Lab: RDD and DataFrame Operations

Introduction to Spark Streaming

Overview of Spark Streaming

Discretized Stream (DStream) operations

Windowed operations and stateful processing

Performance Optimization Basics

Best practices for efficient Spark code

Broadcast variables and accumulators

Monitoring Spark applications

Integrating External Libraries and Tools, Spark Streaming

Using popular external libraries, such as Hadoop and HBase

Integrating with cloud platforms: AWS, Azure, GCP

Connecting to data storage systems: HDFS, S3, Cassandra, etc.

Lab: Building an End-to-End Spark Application: Create a Spark application to process a large dataset, perform aggregations, and save the results.

Lab: Implement a simple Spark Streaming application to process real-time data.

Machine Learning Basics with Spark MLlib

Introduction to Machine Learning Basics

Overview of machine learning

Supervised and unsupervised learning

Common algorithms and use cases

Introduction to Spark MLlib

Overview of Spark MLlib

MLlib's algorithms and utilities

Data preparation and feature extraction

Lab: Data Preparation with Spark MLlib

Linear Regression and Classification

Linear regression algorithm

Logistic regression for classification

Model evaluation and performance metrics

Clustering Algorithms

Overview of clustering algorithms

K-means clustering

Model evaluation and performance metrics

Collaborative Filtering and Recommendation Systems

Overview of recommendation systems

Collaborative filtering techniques

Implementing recommendations with Spark MLlib

Lab: Basic Machine Learning with Spark MLlib

Lab: Implementing a Recommendation System

Graph Processing and Generative AI Technologies

Introduction to Graph Processing

Overview of graph processing

Use cases and applications of graph processing

Graph representations and operations

Introduction to Spark GraphX

Overview of GraphX

Creating and transforming graphs

Graph algorithms in GraphX

Lab: Graph Processing with GraphX

Big Data Innovation! Using GPT and Generative AI Technologies with Spark and Scala

Overview of generative AI technologies

Integrating GPT with Spark and Scala

Practical applications and use cases

Bonus Topics / Time Permitting

Introduction to Spark NLP

Overview of Spark NLP

Preprocessing text data

Text classification and sentiment analysis

Lab: Generative AI Technologies and Spark NLP: Integrate GPT for text generation in a Spark application and explore basic NLP tasks using Spark NLP.

Lab: Text Classification with Spark NLP

Putting It All Together

Work on a capstone project that integrates multiple aspects of the course, including data processing, machine learning, graph processing, and generative AI technologies.

Additional Course Details

Nexus Humans Mastering Scala with Apache Spark for the Modern Data Enterprise (TTSK7520) 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 Mastering Scala with Apache Spark for the Modern Data Enterprise (TTSK7520) 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 Mastering Scala with Apache Spark for the Modern Data Enterprise (TTSK7520)?

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 5-day Mastering Scala with Apache Spark for the Modern Data Enterprise (TTSK7520) course provides up to 32.5 CPD hours of structured learning. CPD certificates can be provided upon request.

Q: What is the duration of the Mastering Scala with Apache Spark for the Modern Data Enterprise (TTSK7520) training?

The training takes place over 5 day(s), with each day lasting approximately 40.00 hours including breaks for lunch and refreshments.

Q: Do you provide corporate training for Mastering Scala with Apache Spark for the Modern Data Enterprise (TTSK7520)?

Yes, we provide corporate training, dedicated training, and closed classes for Mastering Scala with Apache Spark for the Modern Data Enterprise (TTSK7520). 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: Scala; Spark; Data Science

Q: Why choose Nexus Human for Mastering Scala with Apache Spark for the Modern Data Enterprise (TTSK7520)?

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 Mastering Scala with Apache Spark for the Modern Data Enterprise (TTSK7520) 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 13/11/2026.