Apache Spark with Scala Training for Big Data Solutions

Level: Intermediate

RATING: 4.40/5 Based on 78 Reviews

In this hands-on Apache Spark with Scala course you will learn to leverage Spark best practices, develop solutions that run on the Apache Spark platform, and take advantage of Spark's efficient use of memory and powerful programming model. Learn to supercharge your data with Apache Spark, a big data platform well-suited for iterative algorithms required by graph analytics and machine learning.

Key Features of this Apache Spark with Scala Training

- After-course instructor coaching benefit
- Learning Tree end-of-course exam included
- After-course computing sandbox included

You Will Learn How To

- Develop applications with Spark
- Work with the libraries for SQL, Streaming, and Machine Learning
- Map real-world problems to parallel algorithms
- Build business applications that integrate with Spark

Choose the Training Solution That Best Fits Your Individual Needs or Organizational Goals

Live, Instructor-Led - In Class & Live, Online Training

PRODUCT #1262 - $2990

- 4-day instructor-led training course
- After-course instructor coaching benefit
- Learning Tree end-of-course exam included
- Earn 23 NASBA credits (live, in-class training only)

Training At Your Site - Team Training

PRODUCT #1262 - $2990

- Bring this or any training to your organization
- Full - scale program development
- Delivered when, where, and how you want it
- Blended learning models
- Tailored content
- Expert team coaching

In Class & Live, Online Training

Apache Spark with Scala Course Information

- **Requirements**
    - Professional experience in programming at the level of:
        - Course 471, Java Programming Introduction, or
        - Course 419, Introduction to C# Programming, or
        - Course 8476, Programming in C# (20483)
    - Three to six months of experience in a object-oriented programming language

Apache Spark with Scala Course Outline

- **Introduction to Spark**
    - Defining Big Data and Big Computation
    - What is Spark?
    - What are the benefits of Spark?
- **The Challenge of Parallelizing Applications**

    Scaling-out applications

    - Identifying the performance limitations of a modern CPU
    - Scaling traditional parallel processing models

    Designing parallel algorithms

    - Fostering parallelism through functional programming
    - Mapping real-world problems to effective parallel algorithms
- **Defining the Spark Architecture**

    Parallelizing data structures

    - Partitioning data across the cluster using Resilient Distributed Datasets (RDD) and DataFrames
    - Apportioning task execution across multiple nodes
    - Running applications with the Spark execution model

    The anatomy of a Spark cluster

    - Creating resilient and fault-tolerant clusters
    - Achieving scalable distributed storage

    Managing the cluster

    - Monitoring and administering Spark applications
    - Visualizing execution plans and results
- **Developing Spark Applications**

    Selecting the development environment

    - Performing exploratory programming via the Spark shell
    - Building stand-alone Spark applications

    Working with the Spark APIs

    - Programming with Scala and other supported languages
    - Building applications with the core APIs
    - Enriching applications with the bundled libraries
- **Manipulating Structured Data with Spark SQL**

    Querying structured data

    - Processing queries with DataFrames and embedded SQL
    - Extending SQL with User-Defined Functions (UDFs)
    - Exploiting Parquet and JSON formatted data sets

    Integrating with external systems

    - Connecting to databases with JDBC
    - Executing Hive queries in external applications
- **Processing Streaming Data in Spark**

    What is streaming?

    - Implementing sliding window operations
    - Determining state from continuous data
    - Processing simultaneous streams
    - Improving performance and reliability

Streaming data sources

- Streaming from built-in sources (e.g., log files, Twitter sockets, Kinesis, Kafka)
- Developing custom receivers
- Processing with the streaming API and Spark SQL

- **Performing Machine Learning with Spark**

Classifying observations

- Predicting outcomes with supervised learning
- Building a decision tree classifier

Identifying patterns

- Grouping data using unsupervised learning
- Clustering with the k-means method

- **Creating Real-World Applications**

Building Spark-based business applications

- Exposing Spark via a RESTful web service
- Generating Spark-based dashboards

Spark as a service

- Cloud vs. on-premises
- Choosing a service provider (eg, AWS, Azure, Databricks)

- **The Future of Spark**

- Scaling to massive cluster sizes
- Enhancing security on multi-tenant clusters
- Tracking the ongoing commercialization of Spark
- Project Tungsten: pushing performance closer to the limits of modern hardware
- Working with existing projects powered by Spark
- Re-architecting Spark for mobile platforms

Team Training

Apache Spark with Scala Training FAQs

- **What is Scala and Spark?**

  Apache Spark, a big data platform well-suited for iterative algorithms required by graph analytics and machine learning, is written in Scala.

- **Do you need Scala for Spark?**

  Scala is a supported language for Apache Spark. Programming with Scala will help build application with core APIs.

- **Can I learn Apache Spark with Scala online?**

  Yes! We know your busy work schedule may prevent you from getting to one of our classrooms which is why we offer convenient online training to meet your needs wherever you want, including online training.

Schedule of events

In the Classroom — OR — Live, Online

Tuition — Standard: $2990 Government: $2659

Apr 28 - May 1 Greenbelt,MD / Online (AnyWare) (4 Days)

Jun 23 - 26 Herndon, VA / Online (AnyWare) (4 Days)

Aug 4 - 7 Greenbelt,MD / Online (AnyWare) (4 Days)

Sep 8 - 11 New York / Online (AnyWare) (4 Days)

Nov 10 - 13 Greenbelt,MD / Online (AnyWare) (4 Days)

Dec 15 - 18 Herndon, VA / Online (AnyWare) (4 Days)