About Codefrux

While the current trends around the world are based on the internet, mobile and its applications, we try to make the most out of it. As for us, we are a well established IT professionals based in Bangalore, constantly coping up with the extensive advancement and adapting to new Technology. Since the start of the company in 2010, our niches are Mobile Application Development, Web Application Development, Staffing Services and Training Programs (includes Corporate Training).

What you will Learn In This Course

- Hands-on knowledge exploring, running and deploying Apache Spark
- Access to numerous and wide variety of Spark with Scala, Spark SQL, Spark Streaming and Spark MLLib source code examples
- Create hands-on Spark environments for experimenting with course examples
- Participate in course discussion boards with instructor and other students
- Know when and how Spark with Scala, Spark SQL, Spark Streaming and Spark MLLibr may be an appropriate solution

Who can learn Apache Spark and Scala

There is a huge demand for Apache Spark and Scala professionals in IT industry. Big data professionals,Analytics professionals,Research professionals,IT developers and testers,Data scientists,BI and reporting professionals can do this course.

1. **Introduction to Spark**

Topics

a. Overview of BigData and Spark
b. MapReduce limitations
c. Spark History
d. Spark Architecture
e. Limitations of MapReduce in Hadoop Objectives
f. Batch vs. Real-time analytics
g. Application of stream processing
h. Spark and Hadoop Advantages
i. Benefits of Spark and Hadoop
j. Introduction to Spark Eco-system
k. Spark Installation
l. Quiz
m. Summary
n. Hands on

2. **Introduction to Programming in Scala**

Topics

a. Features of Scala
b. Basic data types and literals used
c. List the operators and methods used in Scala
d. Concepts of Scala
e. Scala foundation
f. Features of Scala
g. Setup Spark and Scala on Unbuntu and Windows OS
h. Install IDE's for Scala
i. Run Scala Codes on Scala Shell
j. Understanding Data types in Scala
k. Implementing Lazy Values
l. Control Structures
m. Looping Structures
n. Functions
o. Procedures
p. Collections
q. Arrays and Array Buffers
r. Map's, Tuples and Lists
s. Quiz
t. Summary
u. Hands on

3. OOPS and Functional Programming in Scala

Topics

a. Implementing Classes
b. Getters and Setters
c. Properties with only Getters
d. Object & Object Private Fields
e. Implementing Nested Classes
f. Abstract Classes
g. Constructor
h. Auxiliary and Primary Constructor
i. Singletons
j. Companion Objects
k. Extending a Class
l. Understanding Packages
m. Override Methods
n. Type Checking
o. Casting
p. Overriding Methods
q. Traits as Interfaces
r. Layered Traits
s. Functional Programming
t. Higher Order Functions
u. Anonymous Functions
v. Closures and Currying
w. Performing File Processing
x. Quiz
y. Summary
z. Hands on

4. Foundation to Spark

Topics

a. Spark Shell and PySpark
b. Creating the Spark Context
c. Invoking Spark Shell
d. Loading a file in Shell
e. Basic operations on Shell
f. Spark Java projects
g. Spark Context and Spark Properties
h. Overview of SBT
i. Building a Spark project with SBT
j. Running Spark project with SBT
k. Local mode
l. Spark mode
m. Caching overview
n. Persistance in Spark

o. HDFS data from Spark
p. Implementing Server Log Analysis using Spark
q. Quiz
r. Summary
s. Hands on

5. Working with RDD

Topics

a. Understanding RDD
b. How to create RDDs
c. RDD operations and methods
d. Transformations in RDD
e. Actions in RDD
f. Loading data into RDD
g. Saving data through RDD
h. Key-Value Pair RDD
i. MapReduce and Pair RDD Operations
j. Spark and Hadoop Integration-HDFS
k. Key-Value Pair RDD
l. Scala RDD, Paired RDD, Double RDD & General RDD Functions
m. Implementing HadoopRDD, Filtered RDD, Joined RDD
n. Transformations, Actions and Shared Variables
o. Spark Operations on YARN
p. Sequence File Processing
q. Partitioner and its role in Performance improvement
r. Quiz
s. Summary
t. Hands on

6. Spark Streaming & Spark SQL

Topics

a. Introduction to Spark Streaming
b. Introduction to Spark SQL
c. Querying Files as Tables
d. Text file Format
e. JSON file Format
f. Parquet file Format
g. Hive and Spark SQL Architecture
h. Integrating Spark & Apache Hive
i. Spark SQL performance optimization
j. Implementing Data visualization in Spark
k. Quiz
l. Summary
m. Hands on

7. Spark ML Programming

Topics

a. Explain the use cases and techniques of Machine Learning (ML)
b. Describe the key concepts of Spark ML
c. Explain the concept of an ML Dataset, and ML algorithm, model selection via cross validation
d. Quiz
e. Summary
f. Hands on

8. Spark GraphX Programming

Topics

g. Explain the key concepts of Spark GraphX programming
h. Limitations of the Graph Parallel system
i. Describe the operations with a graph
j. Graph system optimizations
k. Quiz
l. Summary
m. Hands on

Project Work

After course completion, students will be assigned to work on live project to polish the technology skills you have acquired with us.