CREDO SYSTEMZ
Simplifying IT

Apache Spark Course Content

Chapter 1 : Introduction To Scala For Apache Spark

- ➢ What is Scala?
- ➢ Why Scala for Spark?
- ➢ Scala in other frameworks
- ➢ Introduction to Scala REPL
- ➢ Basic Scala operations
- ➢ Variable types in Scala
- ➢ Control Structures in Scala
- ➢ Foreach loop
- ➢ Functions, Procedures, Collections in Scala-Array
- ➢ ArrayBuffer
- ➢ Map, Tuples, Lists, and more.

Chapter 2: Oops And Functional Programming In Scala

- ➢ Class in Scala
- ➢ Getters and Setters
- ➢ Custom Getters and Setters
- ➢ Properties with only Getters
- ➢ Auxiliary Constructor
- ➢ Primary Constructor
- ➢ Singletons
- ➢ Companion Objects
- ➢ Extending a Class
- ➢ Overriding Methods
- ➢ Traits as Interfaces
- ➢ Layered Traits
- ➢ Functional Programming
- ➢ Higher Order Functions
- ➢ Anonymous Functions and more.

Chapter 3: Introduction To Big Data And Apache Spark

➢ Introduction to big data
➢ Challenges with big data
➢ Batch Vs. Real Time big data analytics
➢ Batch Analytics – Hadoop Ecosystem Overview
➢ Real-time Analytics Options
➢ Streaming Data- Spark
➢ In-memory data- Spark
➢ What is Spark?
➢ Spark Ecosystem
➢ Modes of Spark
➢ Spark installation demo
➢ Overview of Spark on a cluster
➢ Spark Standalone cluster o Spark Web UI.

Chapter 4 : Spark Common Operations

➢ Invoking Spark Shell
➢ Creating the Spark Context
➢ Loading a file in Shell
➢ Performing basic Operations on files in Spark Shell
➢ Overview of SBT
➢ Building a Spark project with SBT
➢ Running Spark project with SBT
➢ Local mode
➢ Spark mode
➢ Caching overview
➢ Distributed Persistence

Chapter 5 : Playing With Rdds

➢ RDDs
➢ Transformations in RDD
➢ Actions in RDD
➢ Loading data in RDD
➢ Saving data through RDD
➢ Key-Value Pair RDD
➢ MapReduce and Pair RDD Operations
➢ Spark and Hadoop Integration-HDFS
➢ Handling Sequence Files and Partitioner
➢ Spark Hadoop Integration-Yarn

Chapter 6 : Spark Streaming And Mllib

- Spark Streaming Architecture
- First Spark Streaming Program
- Transformations in Spark Streaming
- Fault tolerance in Spark Streaming
- Checkpointing
- Parallelism level
- Machine learning with Spark
- Data types
- Algorithms– statistics
- Classification and regression
- Clustering
- Collaborative filtering

Chapter 7 : Graphx, Sparksql And Performance Tuning In Spark

- Analyze Hive and Spark SQL architecture
- SQLContext in Spark SQL
- Working with DataFrames
- Implementing an example for Spark SQL
- Integrating hive and Spark SQL
- Support for JSON and Parquet File Formats
- Implement data visualization in Spark
- Loading of data
- Hive queries through Spark
- Testing tips in Scala
- Performance tuning tips in Spark
- Shared variables: Broadcast Variables
- Shared Variables: Accumulators.

Chapter 8: A Complete Project On Apache Spark

- What is a Service?
- Creating a Basic Service
- What is Dependency Injection?
- What Dependency Injection Looks Like
- Injecting Services
- Using a Service in a Component: Dedicated Instance
- Using onInit to Initialize Component Data
- Using a Shared Service Instance
- Dependency Injection
- URL entry and Bookmarking

Contact Info:

📞 **+91 9884412301 | +91 9884312236**

🌐 **Know more about Apache Spark**

✉️ **info@credosystemz.com**

🏢 **New # 30, Old # 16A, Third Main Road, Rajalakshmi Nagar, Velachery, Chennai (Opp. to MuruganKalyanaMandapam)**

BOOK A FREE DEMO

CREDO SYSTEMZ