Big Data Analysis with Scala and Apache Spark

Introduction to Big Data Analysis

- What is Big Data?
- Characteristics of Big Data
- Importance of Big Data Analysis
- Overview of Scala and Spark

Getting Started with Scala

- Introduction to Scala
- Installing Scala
- Scala Basics
    - Variables and Data Types
    - Control Structures
    - Functions and Methods
    - Collections and Data Structures
    - Pattern Matching
    - Exception Handling
    - Object-Oriented Programming in Scala
    - Functional Programming Concepts
    - Type Inference

Introduction to Apache Spark

- Overview of Apache Spark

- Spark Architecture

- Spark Components

    - Spark Core
    - Spark SQL
    - Spark Streaming
    - Spark MLlib
    - Spark GraphX

Working with Spark DataFrames

- Introduction to DataFrames

- Creating DataFrames

- Reading and Writing Data

- Transformations on DataFrames

- Actions on DataFrames

- Working with Columns and Functions

- Filtering and Sorting Data

Spark SQL and Data Processing

- Introduction to Spark SQL

- Working with SQL Queries

- DataFrame API vs. SQL

- Aggregations and Grouping

- Joins and Broadcast Joins

- Window Functions

- Working with User-Defined Functions (UDFs)

Spark Streaming

- Introduction to Spark Streaming

- DStreams (Discretized Streams)

- Streaming Transformations

- Windowed Operations

- Stateful Operations

- Integrating Spark Streaming with other systems

Machine Learning with Spark MLlib

- Introduction to Spark MLlib
- ML Pipelines
- Data Preparation and Feature Engineering
- Model Building and Evaluation
- Regression
- Classification
- Clustering
- Recommendation Systems

Graph Processing with Spark GraphX

- Introduction to Graph Processing
- GraphX API
- Creating Graphs
- Graph Operations
- Graph Algorithms

Performance Optimization and Tuning

- Understanding Spark Performance
- Identifying Performance Bottlenecks
- Data Partitioning and Storage Formats
- Caching and Persistence
- Spark Configuration and Tuning

Real-Time Big Data Processing

- Introduction to Real-Time Processing
- Stream Processing with Spark Streaming
- Stream Processing with Structured Streaming
- Integrating Spark with Apache Kafka

- Handling Streaming Data with Spark

Advanced Topics in Spark

- Working with Spark on a Cluster

- Spark Ecosystem and Integration with other Tools

- Spark and Hadoop Integration

- Spark and Hive Integration

- Spark and Cassandra Integration

- Spark and Elasticsearch Integration

Scala and Spark Integration

- Scala and Spark: Key Similarities and Differences

- Scala Collections and Spark RDDs

- Scala Higher-Order Functions and Spark

- Working with Scala and Spark together in a Project