London TDM

Artificial Intelligence and Data Science Training Courses

Course Venue: Malaysia - Kuala Lumpur

Course Date: From 01 February 2026 To 05 February 2026

Course Place: Royale Chullan Hotel

Course Fees: 6,000 USD

Introduction

This comprehensive 5-day course on "Big Data Analytics with Hadoop and Spark" is designed to equip professionals with the necessary skills to harness the power of big data technologies. Participants will delve into the fundamentals of Hadoop and Spark, understand their architecture, and learn how to implement real-world big data solutions effectively.

- Understand the core concepts of big data and the significance of Hadoop and Spark.
- Gain hands-on experience with Hadoop ecosystems and Spark frameworks.
- Explore data processing, storage, and analytics using Hadoop and Spark.
- Develop skills to optimize performance and troubleshoot big data applications.
- Acquire knowledge to implement big data solutions in various sectors.

Course Outlines

Day 1: Introduction to Big Data and Hadoop Ecosystem

- Overview of Big Data and its characteristics
- Introduction to Hadoop and its architecture
- Understanding HDFS (Hadoop Distributed File System)
- Introduction to MapReduce programming model
- Overview of Hadoop ecosystem components: Hive, Pig, HBase, etc.

Day 2: Hadoop Setup and MapReduce Programming

- Setting up Hadoop environment and cluster configuration
- Basics of MapReduce and its data flow
- Writing a MapReduce Job in Java/Python
- Understanding Combiner, Partitioner, and Reducer
- Troubleshooting and debugging MapReduce jobs

Day 3: Introduction to Apache Spark

- Understanding Spark architecture and its components
- Setting up Spark environment
- Introduction to RDDs (Resilient Distributed Datasets)
- Basic transformations and actions in Spark
- Exploring Spark SQL and DataFrames

Day 4: Advanced Spark Programming and Optimization

- Understanding Spark streaming and real-time data processing
- Introduction to Spark MLlib and machine learning with Spark
- Optimizing Spark applications with partitioning and caching
- Tuning Spark performance and job scheduling
- Troubleshooting and debugging Spark applications

Day 5: Implementing Big Data Solutions

- Designing end-to-end big data solutions with Hadoop and Spark
- Integrating Hadoop and Spark with other tools (e.g., Kafka, Flume)
- Case studies and real-world big data applications
- Hands-on project: Building a big data analytics application
- Course review and Q&A session