# Big Data Engineering for Analytics: Architecting Solutions

2027 □□□□ 12 – 8
Kuala Lumpur (Malaysia)



# Big Data Engineering for Analytics: Architecting Solutions

Ref: 20220289_138018 Duration: Kuala Lumpur (Malaysia) Fees: 4900 GBP

## Why This Course

In the era of digital transformation, the ability to harness, process, and analyze massive volumes of data has become a critical competitive advantage. The Big Data Engineering and Analytics Program is an intensive 1-week course designed to equip data professionals, engineers, and analysts with the practical skills and architectural understanding required to design, implement, and manage large-scale data systems for analytics and machine learning.

Through a balanced combination of technical theory, hands-on labs, and case studies, participants will master the tools and frameworks that power modern data ecosystems — including data lakes, distributed computing platforms, and real-time data processing pipelines. This program prepares professionals to build scalable, secure, and high-performance data infrastructure for enterprise and cloud environments.

## What You'll Learn and Practice

By completing this program, participants will:

- Design and deploy scalable big data architectures tailored for analytics.
- Build and manage data lakes using distributed storage and management systems.
- Develop end-to-end batch and streaming data pipelines using modern frameworks.
- Apply best practices for governance, data quality, and security in large-scale systems.
- Integrate big data platforms with analytics and machine learning workflows.

## The Program Flow

### Day 1: Introduction to Big Data Engineering

- Big data concepts, evolution, and ecosystem overview.
- Fundamentals of distributed systems and parallel data processing.
- Big data architecture design patterns and reference models.
- Real-world use cases and success stories in big data applications.

### Day 2: Data Storage and Management

- Designing and implementing scalable data lakes.
- Distributed file systems: HDFS, Amazon S3, and object storage.
- NoSQL databases Cassandra, MongoDB for unstructured and semi-structured data.
- Data modeling, partitioning, and schema design for big data environments.



### Day 3: Data Processing and Analytics

- Batch data processing with Hadoop and MapReduce.
- Stream processing using Apache Kafka, Spark Streaming, and Flink.
- SQL on big data with Apache Hive, Impala, and Presto.
- Scalable machine learning and analytics using Spark MLlib and TensorFlow.

### Day 4: Data Pipelines and Workflow Management

- ETL design principles and integration for big data systems.
- Workflow orchestration with Apache Airflow and other automation tools.
- Data validation, lineage tracking, and quality management.
- Monitoring, alerting, and optimizing data pipelines for reliability.

### Day 5: Advanced Topics and Best Practices

- Data governance frameworks and compliance strategies.
- Security and access control in big data platforms.
- Performance tuning, resource optimization, and cost management.
- Real-time analytics, visualization, and dashboarding solutions.
- Capstone case studies and industry applications.

## Individual Impact

- Gain hands-on expertise in designing and managing big data infrastructure.
- Learn to build efficient and reliable data pipelines for analytics and AI.
- Strengthen technical skills in distributed computing and real-time data processing.
- Enhance professional value through mastery of modern big data technologies.

## Organizational Impact

- Build scalable, high-performance data systems for advanced analytics.
- Improve data quality, accessibility, and decision-making across business units.
- Reduce operational complexity through automation and optimized data pipelines.
- Accelerate innovation and insight generation through modern data architecture.

## Training Methodology

This program combines conceptual depth with applied practice through:

- Instructor-led sessions on big data architecture and design.
- Hands-on exercises using industry tools such as Hadoop, Spark, and Airflow.
- Group workshops on data pipeline development and optimization.
- Case studies illustrating successful enterprise big data implementations.



## Beyond the Course

Upon completion, participants will be equipped to design and manage large-scale, analytics-driven data environments capable of supporting modern AI and business intelligence initiatives.

**Graduates of this program will emerge as skilled data engineers and analytics innovators — ready to lead digital transformation through data-driven excellence.**

