

# Big Data Engineering for Analytics

IT & IT Engineering  
Cairo (Egypt)  
02 - 06 Feb 2025

UK Training

# PARTNER



## Big Data Engineering for Analytics

**Ref:** 3289\_127842 **Date:** 02 - 06 Feb 2025 **Location:** Cairo (Egypt) **Fees:** 3300 **Euro**

### Introduction

This course helps data engineers focus on essential design and architecture while building a data lake and relevant processing platform. Participants will learn various aspects of data engineering while building resilient distributed datasets *RDDs*. Participants will learn to apply key data engineering practices, identify multiple data sources appraised against their business value, design the right storage, and implement proper access models.

### Course Objectives of Big Data Engineering for Analytics

- Understand the fundamental characteristics, storage, analysis techniques, and the relevant distributions.
- Gain expertise with the fault-tolerant computing framework within the Hadoop ecosystem in big data.
- Construct configurable and executable tasks within big data processing systems.
- Understand the nuances of writing functional programs in big data engineering.
- Understand various data processing, querying, and persistence available in RDDs through spark rdd and rdd in spark.
- Perform filtering, selection, and categorization tasks in data engineering initiatives.

### Course Outlines of Big Data Engineering for Analytics

#### Day 1: Data Science, Data Engineering, Big Data, and Analytics Perspective

- Introduction to Data Science, Data Engineering, and Big Data.
- Data Scientist vs. Data Engineer.
- Different Roles in Data Engineering.
- Core Data Engineering Skills and Resources.
- Understand Big Data from an Analytics Perspective.

#### Day 2: Architectural Viewpoints and Hadoop Ecosystem

- Architectural Viewpoints in Big Data.
- Reference Architecture Conceptual View.
- Reference Architecture Logical View.
- Oracle Product Mapping View.
- The Hadoop Ecosystem for Big Data.

A graphic of a chessboard with several chess pieces (a king, a queen, a rook, and a pawn) on it. The pieces are gold and silver. The board is white and black squares. The text 'UK Training PARTNER' is overlaid on the right side of the board.

UK Training  
**PARTNER**

### Day 3: File Storage and Databases for Big Data

- Distributed File Storage.
- NoSQL Databases for Big Data.
- Spark and Functional Programming for Big Data.

### Day 4: Management of Big Data

- Spark and Resilient Distributed Data Sets.
- Spark QL for Big Data.
- Spark and Real-Time Stream Processing.
- Management of Big Data initiatives.

### Day 5: Dealing with a case study

- Case study.
- Project Requirement Elaboration.
- Project and Assessment.
- Project Demonstration.
- Report Submission and Presentations.

## Data Lake Design and Data Engineering Best Practices

Within the realm of *Big Data Engineering*, it's critical to understand how to design and maintain a data lake that is both scalable and manageable. Participants will learn how to architect a data lake that adheres to the best data engineering practices, ensuring that it can handle the influx of data from varied sources while maintaining high performance for big data processing.

## Fault-Tolerant Computing and the Hadoop Ecosystem

The course will focus on fault-tolerant computing within the context of the Hadoop ecosystem. Enabling systems to continue operating properly in the event of the failure of some of their components is a cornerstone of big data engineering. This part of the course will explore how the Hadoop ecosystem provides fault-tolerant mechanisms through its various components.

## Spark RDDs in Big Data Engineering

Understanding Resilient Distributed Datasets RDDs and their role in big data processing will be crucial. This course will cover how RDDs within the Apache Spark framework provide a fault-tolerant way to work with large datasets across multiple nodes in a cluster. Participants will gain hands-on experience with spark rdd and rdd in spark, implementing complex data transformations and actions in a distributed environment.

A graphic of a chessboard with several chess pieces (a king, a queen, a rook, and a knight) on it. The board is white and black, and the pieces are gold and silver. The text 'UK Training PARTNER' is overlaid on the board.

UK Training  
**PARTNER**

## Blackbird training cities

Accra1 (Ghana)

Amman (Jordan)

Amsterdam (Netherlands)

Annecy (France)

Baku (Azerbaijan)

Bali (Indonesia)

Bangkok (Thailand)

Bangkok (Thailand)

Barcelona (Spain)

Batumi (Georgia)

Beijing (China)

Beirut (Lebanon)

Berlin (Germany)

Birmingham (UK)

Bordeaux (France)

Boston,Massachusetts (USA)

Brussels (Belgium)

Cairo (Egypt)

Cape Town (South Africa)

Casablanca (Morocco)

Cascais (Portugal)

Copenhagen (Denmark)

Doha (Qatar)

Dubai (UAE)

Düsseldorf (Germany)

UK Traininig  
**PARTNER**



## Blackbird Training Category



Human Resources



Audit & Quality Assurance



Finance, Accounting, Budgeting



Marketing, Sales, Customer Service



Secretary & Admin



Law and Contract Management



Project Management



IT & IT Engineering



Supply Chain & Logistics



Management & Leadership



Professional Skills



Oil & Gas Engineering



Health & Safety



Telecom Engineering



Hospital Management



Customs & Safety



Aviation



C-Suite Training



Agile and Refinement



## Blackbird training Clients



UK Training  
**PARTNER**



**BLACKBIRD**  
FOR TRAINING

**LONDON TRAINING PROVIDER**



[www.blackbird-training.com](http://www.blackbird-training.com)



[training@blackbird-training.com](mailto:training@blackbird-training.com)



+44 7480 775526 / +44 7401 177335