

Big Data Tools: Hadoop, Spark & NoSQL

COURSE OVERVIEW

This practical course provides a comprehensive introduction to the core tools and technologies that form the foundation of modern big data ecosystems. The content focuses on understanding the architecture, use cases, and practical implementation of these technologies to process and analyze large-scale datasets that exceed the capabilities of traditional systems, preparing you to design and work with scalable data solutions. Participants will acquire practical experience with the Hadoop Distributed File System (HDFS) for storage, MapReduce and Spark for distributed processing, and popular NoSQL databases like MongoDB and HBase for handling unstructured data.

WHO SHOULD ATTEND?

This course is designed for data engineers, software developers, database administrators, data analysts, and IT professionals who need to process, store, and manage large volumes of data using distributed computing frameworks and non-relational database technologies.

COURSE OUTCOMES

Delegates will gain the skills and knowledge to:

- Understand the architecture and components of Hadoop and Spark ecosystems.
- Process large datasets using MapReduce and Spark transformations.
- Work with HDFS for distributed data storage.
- Implement basic operations using NoSQL databases.
- Choose appropriate big data tools for different use cases.
- Develop basic data pipelines using big data technologies.

KEY COURSE HIGHLIGHTS

At the end of the course, you will understand;

- The Hadoop ecosystem architecture including HDFS, YARN, and MapReduce.
- Spark's in-memory computing advantages over traditional MapReduce.
- How to process structured and unstructured data using Spark RDDs and DataFrames.
- The CAP theorem and its implications for database design.
- Different types of NoSQL databases and their appropriate use cases.
- How to integrate multiple big data tools into complete data processing pipelines.

All our courses are dual-certificate courses. At the end of the training, the delegates will receive two certificates.

1. A GTC end-of-course certificate
2. Continuing Professional Development (CPD) Certificate of completion with earned credits awarded