Big Data Technology Fundamentals

Who is the course for

Individuals who are new to big data concepts, including Enterprise Solutions Architects, Big Data Solutions Architects,

Data Scientists, and Data Analysts.

What we teach you

This course teaches you how to:

- Identify common tools and technologies that can be used to create big data solutions
- Understand the MapReduce programming framework, including the map, shuffle and sort, and reduce components
- Distinguish options available for creating a big data solution using the Hive programming framework

Please register for free here.

Required skills

We recommend that attendees of this course have:

- Working knowledge of basic programming in a language such as Java or C#

Course outline

Module 1 – Introduction to Big Data

- The Business Importance of Big Data
- The Hadoop Ecosystem
- Characteristics of Big Data
- Processing Big Data
- Tools and Techniques for Analyzing Big Data
- Implementing Big Data Solutions
- Case Study – Social Media Analytics

Module 2 – Introduction to MapReduce and Hadoop

- Hadoop Architecture
- MapReduce Framework
- MapReduce Programming

GOPAS®

- MapReduce and HDFS/S3
- Use Case – Recommendation Engine

Module 3 – Data Analysis Using Pig Programming
- Introduction to Pig
- Pig Data Types
- Representing Data in Pig
- Running Pig
- User-Defined Functions
- Pig vs Traditional RDBMSs
- Advanced Techniques in Pig

Module 4 – Big Data Querying with Hive
- Introduction to Hive
- Representing Data in Hive
- Hive Data Types
- Probing Data with Hive Queries
- Hive and AWS
- Use Case – Ad Hoc Analysis and Product Feedback