

Certified Big Data and Data Analytics Practitioner

Course Description:

This comprehensive certification course is designed to provide participants with an in-depth understanding of Big Data Analytics, covering essential concepts, technologies, and practices. Participants will learn how to effectively manage, store, and analyze large data sets to derive actionable insights and drive business decisions.

Duration: 5 Days

Pre-requisites:

- Basic understanding of data analytics concepts
- Familiarity with programming languages such as Python or Java
- Basic knowledge of databases and SQL
- Prior experience with data analysis tools and techniques is recommended

Course Content:

Module 1: Introduction to Big Data Analytics

- What is Big Data?
- The 5 “V’s” of Big Data

Module 2: Big Data in Context

- How Big Data relates to Data Analytics
- Impact of Big Data on Technologies
- The Open Source Revolution

Module 3: Key Big Data Concepts and Data Types

- Types of Data: Text, Audio, Images
- Professional Roles in Big Data

Module 4: Big Data Projects and Organizational Impact

- Meeting Organizational Needs with Big Data
- Examples of Big Data in Use: Netflix, LinkedIn, Facebook, Google, Orbitz, Dell, Others

Module 5: Best Practices in Project Design

- Assessing the Current State of Your Organization

Module 6: Storing Big Data

- Big Data Architectures and Paradigms
- The Hadoop Ecosystem:
 - Overview of Hadoop
 - Hadoop Distributed File System (HDFS)
- Massively Parallel Processing (MPP) vs Distributed In-Memory Applications
- RDBMSs vs NoSQL DBs:
 - PostgreSQL, MongoDB, Cassandra
- Streaming Data
- Data-Warehousing vs Data Mart

Module 7: Computing Big Data

- Accessing Big Data
- The Role of Cloud Computing
- Data Movement Risks
- Networking and Co-Location
- Big Data Extract, Transform, Load (ETL)
- Big Data Compute Technologies:
 - Hadoop Continued: MapReduce and Beyond
 - Distributed Compute
 - High Performance Clusters
 - Spark
 - Streaming Technologies: Storm, Spark Structured Streaming
 - Other Big Data Technologies: Kafka, etc.
 - Cloud Applications for Big Data

Module 8: Big Data Projects Implementation

- Basics of Data Analytics:
 - Roles and Objectives
 - Key Math and Statistics Concepts
 - Supervised vs Unsupervised Learning
 - Key Technologies and Applications

Module 9: Extracting Value from Big Data

- The 5 P's of Data Science
- The Importance of Ethics
- Programmability

Module 10: Architecting Big Data Solutions

- Identifying Analytical Opportunities
- Defining and Assessing Problems
- Describing Impact and Use of Data
- Identifying Potential Data Sources
- Brainstorming an Analytics Strategy to Implement

Module 11: Storage and Compute Strategy

- Identifying a Cloud Environment Strategy
- Brainstorming Key Storage Systems and Compute Environments