

Certified Big Data and Data Analytics Practitioner

Why Attend

- Big data is a change agent that challenges the ways in which organizational leaders have traditionally made decisions. This course provides participants with the confidence to articulate big data architectures to support analytics driven solutions within their organizations. This course provides hands on experience with key big data technologies used to solve data intensive problems. Participants will gain the knowledge and skills they need to assemble and manage a large-scale big data analytics project. Lastly, participants will receive a conceptual introduction to the data structures that support machine learning algorithms and artificial intelligence use cases.
- Participants will work to identify areas within their organization that can be improved through big data-driven implementations, and the types of improvements that can be made through analytical processes. Participants will be led through a series of hands-on exercises and workshops, where they will have the chance to apply the test methods and practical approaches that they are learning throughout the course. At the end of the course, participants will produce an actionable big data plan and architectural diagram to be used as a blueprint proposal within their own organization.

Course Methodology

- This course will be highly interactive with group discussions, case studies, hands-on practical exercises, and group activities being the core focus.

Course Objectives

By the end of the course, participants will be able to:

- Design big data implementation plans and create strategies for data driven solutions
- Explain the challenges of big data and traditional technologies like Excel
- Discuss the main challenges and advantages of Hadoop ecosystem and other big data distributed architectures
- Demonstrate and discuss key technologies for big data storage and compute, such as PostgreSQL and MongoDB
- Discuss popular machine learning algorithms and the importance of ethics in data analytics and artificial intelligence

Target Audience

- This course is ideal for data analysts, data engineers, data scientists, as well as technically-inclined management and administrative professionals seeking to understand big data strategies, technologies and use cases. Recommended pre-knowledge includes basic programming experience and analyzing data in python, knowledge of basic database technologies, and awareness of analytics driven business initiatives.

Target Competencies

- Big data hands-on labs
- Big Data analytics structures and technologies
- Ethics and integrity for big data analytics
- Big data storage and computer system implementation
- Architecture diagram design

Introduction to Big Data Analytics

- What is Big Data?
- 5 “V’s” of big data
- How big data relates to data analytics
- Big data impact on technologies
- Open source revolution
- Key big data concepts and data types
- Text, audio, images
- Big data professional roles
- How can big data projects meet organizational needs
- Big data Examples:
- Netflix, LinkedIn, Facebook, Google, Orbitz, Dell, others.
- Best practices in project design
- Assessing the current state of your organization

Storing Big Data

- Big data architectures and paradigms
- The Hadoop Ecosystem
- Overview of Hadoop
- Hadoop Distributed File System (HDFS)
- Massively parallel processing (MPP) versus distributed in-memory applications
- RDBMSs vs NoSQL DBs
- PostgreSQL, MongoDB, Cassandra
- Streaming data
- Data-warehousing versus Data Mart

Computing Big Data

- How to access big data
- Role of cloud computing
- Data movement risk
- 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: Storm, Spark structured streaming
- Other big data technologies: Kafka, etc.
- Cloud applications for big data

Big Data Projects

- Basics of data analytics
- Roles and objectives
- Key math and statistics concepts
- Supervised versus Unsupervised
- Key technologies and applications
- Getting value out of Big Data
- 5 P's of data science
- Importance of ethics
- Programmability

Architecting Big Data Solutions

- Identify analytical opportunities
- Define and assess the problem
- Describe the impact and use of data to address the problem
- Identify potential data sources
- Brainstorm an analytics strategy to implement
- Storage and compute
- Identify a cloud environment strategy

Brainstorm key storage systems and compute environments