

€ TRAINING

Big Data and Data Analytics



Big Data and Data Analytics

Introduction:

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. The course also provides hands-on experience with key big data technologies used to deploy data-intensive applications. 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 opportunity to apply the test methods and practical approaches that they learn 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 organizations.

Course Objectives:

At the end of this course the 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 the 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
- Deliver an architectural diagram for analytics-focused use cases

Targeted Audience:

This course is ideal for data professionals, such as database administrators, system administrators, business analysts, or business intelligence specialists. It is also ideal for less technically-inclined management and administrative professionals seeking to understand big data strategies and technologies.

Course Outlines:

Unit 1: Storing Big Data:

- What is big data?
- 5 Vs 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
- 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 vs Data Mart
- Lambda Architecture vs Kappa Architecture

Unit 2: 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

Unit 3: Introducing Big Data Analytics and Artificial Intelligence AI:

- Basics of data analytics
- Roles and objectives
- Key math and statistics concepts
- Supervised vs Unsupervised
- Key technologies and applications
- Analytics architecture
- Cloud vs On-premise
- Data storage
- Analytics Tools
- Databricks
- SAS Viya
- Cloud ML & AI solutions
- Introduction to Artificial Intelligence
- Linear Algebra 101
- Image classification
- Importance of Ethics

Unit 4: Planning A Big Data Project For Analytics:

- How big data projects meet organizational needs
- Big data case studies:

- Netflix
- LinkedIn
- Facebook
- Google
- Orbitz
- Dell
- And others
- Best practices in project design
- Assessing the current state of your organization
- Vertical data teams and discussions
- Considerations for big data project plans
- Brainstorm a data-driven strategy
- Practice designing architecture diagrams

Unit 5: Architecting Big Data Solutions:

- Identifying 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