



Data Science & Bigdata Analytics

Overview

This course provides practical foundation level training that enables immediate and effective participation in Big Data and other analytics projects. It includes an introduction to Big Data and the data analytics lifecycle to address business challenges that leverage Big Data.

The course provides grounding in basic and advanced analytic methods and an introduction to Big Data analytics technology and tools, including MapReduce and Hadoop. Labs offer opportunities for students to understand how these methods and tools may be applied to real world business challenges by a practicing data scientist.

The course takes an “open”, or technology-neutral approach and includes a final lab which addresses a big data analytics challenge by applying the concepts taught in the course in the context of the data analytics lifecycle.

The course prepares the student for the Dell EMC Proven™ Professional Data Scientist Associate (DECA-DS) certification exam.

Who Should Attend?

- Managers of teams of business intelligence, analytics, and big data professionals
- Current Business and Data Analysts looking to add big data analytics to their skills.
- Data and database professionals

Course Outcome

Upon successful completion of this course, participants should be able to:

- Immediately participate as a data science team member
- Work with large data sets and generate insights
- Build predictive and classification models
- Manage a data analytics project through the entire lifecycle

Duration

5 days

Certificate

Dell EMC Proven™ Professional Data Scientist Associate (DECA-DS)

Certification Body

Dell Technologies

Course Outline

The content of this course is designed to support the course objectives.

- Module 1 - Introduction to Big Data analytics
 - Big Data and its characteristics Lesson
 - Business value from Big Data
 - Data scientist
- Module 2 – Data Analytics Life Cycle
 - Data analytics life cycle overview
 - Discovery phase
 - Data preparation phase
 - Model planning phase
 - Model building phase
 - Communicate results phase
 - Operationalize phase
- Module 3 – Basic data analytics methods using R
 - Introduction to the R programming language
 - Analyzing and exploring data
 - Statistics for model building and evaluation
- Module 4– Advanced analytics theory and methods
 - Introduction to advanced analytics—theory and methods
 - Association rules
 - Linear regression
 - Logistic regression
 - Text analysis
 - Naïve Bayes
 - Decision trees
 - Time series analysis
- Module 5: Advanced analytics—technology and tools
 - Introduction to advanced analytics—technology and tools
 - Hadoop ecosystem
 - In-database analytics SQL essentials
 - Advanced SQL and MADlib
- Module 6: Putting it all together
 - Preparing to operationalize
 - Preparing project presentations
 - Data visualization techniques