

EMC - Data Science and Big Data Analytics v2

Code:	4447
Length:	5 days
URL:	View Online

In this course, you will gain practical foundation level training that enables immediate and effective participation in big data and other analytics projects. You will cover basic and advanced analytic methods and big data analytics technology and tools, including MapReduce and Hadoop. Extensive labs throughout the course provide you with the opportunity to apply these methods and tools to real world business challenges using a technology-neutral approach. In a final lab, you will address a big data analytics challenge by applying the concepts taught in the course to the context of the Data Analytics Lifecycle. You will prepare for the Data Scientist Associate (EMCDSA) certification exam and establish a baseline of Data Science skills.

Suggested Audience

Skills Gained

- Deploy the Data Analytics Lifecycle to address big data analytics projects
- Reframe a business challenge as an analytics challenge
- Apply appropriate analytic techniques and tools to analyze big data, create statistical models, and identify insights that can lead to actionable results
- Select appropriate data visualizations to clearly communicate analytic insights to business sponsors and analytic audiences
- Use R and RStudio, MapReduce/Hadoop, in-database analytics, Windows, and MADlib functions
- Use advanced analytics to create a competitive advantage
- Data scientist role and skills vs. traditional business intelligence analyst

Who Can Benefit

- Managers of business intelligence, analytics, and big data professionals teams
- Current business and data analysts looking to add big data analytics to their skills
- Data and database professionals looking to exploit their analytic skills in a big data environment
- Recent college graduates and graduate students with academic experience in a related discipline looking to move into the world of Data Science and big data
- Individuals looking to take the Data Scientist Associate (EMCDSA) certification

Prerequisites

To successfully complete this course and gain the maximum benefits from participation, you should have the following knowledge and skill sets:

- A strong quantitative background with a solid understanding of basic statistics, as would be found in a statistics 101 level course

- Experience with a scripting language such as Java, Perl, or Python (or R). Many of the lab examples taught in the course use R (with an RStudio GUI), which is an open source statistical tool and programming
- Experience with SQL

Course Details

1. Big Data Analytics

- Big Data
- State of the Practice in Analytics
- Data Scientist
- Big Data Analytics in Industry Verticals

2. Data Analytics Lifecycle

- Discovery
- Data Preparation
- Model Planning
- Model Building
- Communicating Results
- Operationalizing

3. Basic Data Analytic Methods Using R

- Using R to Look at Data
- Analyzing and Exploring the Data
- Statistics for Model Building and Evaluation

4. Advanced Analytics: Theory and Methods

- K Means Clustering
- Association Rules
- Linear Regression
- Logistic Regression
- Naive Bayesian Classifier
- Decision Trees
- Time Series Analysis
- Text Analysis

5. Advanced Analytics: Technologies and Tools

- Analytics for Unstructured Data
 - MapReduce and Hadoop

- Hadoop Ecosystem
 - In-Database Analytics: SQL Essentials
 - Advanced SQL and MADlib for In-Database Analytics

6. Putting it All Together

2012-09-05 13:51:04.617000000

ExitCertified® Corporation and iMVP® are registered trademarks of ExitCertified ULC and ExitCertified Corporation and Tech Data Corporation, respectively
Copyright ©2019 Tech Data Corporation and ExitCertified ULC & ExitCertified Corporation.
All Rights Reserved.

Generated 5