EMC Data Science and Big Data Analytics

Duration: 5 Days   Course Code: DSBDA

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 as a practicing data scientist. The course takes an "Open", or technology-neutral approach, and includes a final lab in which students address 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 Proven™ Professional Data Scientist Associate (EMCDSA) certification exam.

Target Audience:

This course is intended for individuals seeking to develop an understanding of Data Science from the perspective of a practicing Data Scientist.

Objectives:

- Immediately participate and contribute as a Data Science Team Member on big data and other analytics projects by:

- Deploying the Data Analytics Lifecycle to address big data analytics projects

- Reframing a business challenge as an analytics challenge

- Applying appropriate analytic techniques and tools to analyze big data, create statistical models, and identify insights that can lead to actionable results

- Selecting appropriate data visualizations to clearly communicate analytic insights to business sponsors and analytic audiences

- Using tools such as: R and RStudio, MapReduce/Hadoop, in-database analytics, Window and MADlib functions

- Explain how advanced analytics can be leveraged to create competitive advantage and how the data scientist role and skills differ from those of a traditional business intelligence analyst

Prerequisites:

To complete this course successfully and gain the maximum benefits from it, a student 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 (some course examples use

Consider the above as a list of specific prerequisite (or refresher) training and reading to be completed prior to enrolling for or attending this course. Having this requisite background will help ensure a positive experience in the class, and enable students to build on their expertise to learn many of the more advanced tools and analytical methods taught in the course.

Content:

Introduction and Course Agenda

Introduction to Big Data Analytics

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

Data Analytics Lifecycle

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

Review of Basic Data Analytic Methods Using R

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

Advanced Analytics – Theory And Methods

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

Advanced Analytics - Technologies and Tools

- Analytics for Unstructured Data - MapReduce and Hadoop
- The Hadoop Ecosystem
- In-database Analytics – SQL Essentials
- Advanced SQL and MADlib for In-database Analytics

The Endgame, or Putting it All Together

- Operationalizing an Analytics Project
- Creating the Final Deliverables
- Data Visualization Techniques
- Final Lab Exercise on Big Data Analytics

Further Information:

For More information, or to book your course, please call us on 00 20 (0) 2 2269 1982 or 16142

training@globalknowledge.com.eg

www.globalknowledge.com/en-eg/

Global Knowledge, 16 Moustafa Refaat St. Block 1137, Sheraton Buildings, Heliopolis, Cairo