COURSE DESCRIPTION

This course will prepare you to immediately participate as a Data Science team member on big data and data analytics projects. You will learn concepts and skills that are essential for managers of data analytics and business intelligence teams, as well as for database, data warehouse, and big data professionals. Upon the successful completion of this course, you will be able to:

- Translate a business challenge into an analytics challenge;
- Deploy a structured lifecycle approach to data science and big data analytics projects;
- Analyze big data, create statistical models, and identify insights that can lead to actionable results;
- Use software tools such as R and Hadoop, in-database analytics, and window and MADlib functions;
- Select visualization techniques, communicate analytic insights to business sponsors, and others;
- Explain how advanced analytics can be leveraged to create competitive advantage;
- Define and distinguish a data scientist from a traditional business intelligence analyst.

Recommended (but not required) prerequisites include basic quantitative skills, including elementary statistics, as well as basic programming skills in SQL and R (or a similar environment such as Minitab, Matlab, SAS, or SPSS). The course provides hands-on labs for you to gain practical experience in all areas listed above.

MODULES AND TOPICS

- **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
- **Review of the Basic Data Analytic Methods using R**
  - Introduction to R – look at the data
  - Analyzing and Exploring the Data
  - Statistics for Model Building and Evaluation
- **Advanced Analytics**
  - K-means clustering
  - Association rules
  - Linear Regression
  - Logistic Regression
  - Naïve Bayes
  - Decision Trees
  - Time Series Analysis
  - Text Analysis
- **Advanced Analytics**
  - Analytics for Unstructured Data (MapReduce and Hadoop)
  - The Hadoop Ecosystem
  - In-database Analytics – SQL Essentials
  - Advanced SQL and MADlib for in-database Analytics
- **Putting All Together**
  - Operationalizing an Analytics Project
  - Creating the Final Deliverables
  - Data Visualization Techniques
  - Final Lab: Application of Data Analytics Lifecycle to a Big Data Analytics Challenge