ADVANCED METHODS IN DATA SCIENCE AND BIG DATA ANALYTICS

Course Code: 4192

Learn to use several open source tools to address big data challenges.

This course builds on skills developed in the Data Science and Big Data Analytics course. The main focus areas cover Hadoop (including Pig, Hive, and HBase), natural language processing, social network analysis, simulation, random forests, multinomial logistic regression, and data visualization. With a technology-neutral approach, this course utilizes several open-source tools to address big data challenges.

What You’ll Learn

• MapReduce functionality
• NoSQL databases and Hadoop Ecosystem tools for analyzing large-scale, unstructured data sets
• Natural language processing, social network analysis, and data visualization concepts
• Use advanced quantitative methods, and apply one of them in a Hadoop environment
• Apply advanced techniques to real-world datasets in a final lab

Who Needs to Attend

• Aspiring data scientists
• Data analysts that have completed the associate level Data Science and Big Data Analytics course
• Computer scientists wanting to learn MapReduce and methods for analyzing unstructured data such as text.
ADVANCED METHODS IN DATA SCIENCE AND BIG DATA ANALYTICS

Course Code: 4192

Classroom Live Outline

1. MapReduce and Hadoop
   • The MapReduce Framework
   • Apache Hadoop
   • Hadoop Distributed File System
   • YARN

2. Hadoop Ecosystem and NoSQL
   • Hadoop Ecosystem
   • Pig
   • Hive
   • NoSQL--Not only SQL
   • HBase
   • Spark

3. Natural Language Processing
   • Introduction to NLP
   • Text Preprocessing
   • TFIDF
   • Beyond Bag of Words
   • Language Modeling
   • POS Tagging and HMM
   • Sentiment Analysis and Topic Modeling

4. Social Network Analysis
   • Introduction to SNA and Graph Theory
   • Most Important Nodes
   • Communities and Small World
   • Network Problems and SNA Tools
5. Data Science Theory and Methods
   - Simulation
   - Random Forests
   - Multinomial Logistic Regression

6. Data Visualization
   - Perception and Visualization
   - Visualization of Multivariate Data

Classroom Live Labs
In addition to lecture and demonstrations, this course includes labs designed to give you practical experience.
ADVANCED METHODS IN DATA SCIENCE AND BIG DATA ANALYTICS

Course Code: 4192

PRIVATE GROUP TRAINING 5 days

Visit us at www.globalknowledge.com or call us at 1-866-716-6688.

Date created: 1/1/2020 7:37:07 PM
Copyright © 2020 Global Knowledge Training LLC. All Rights Reserved.