

BDSCP MODULE 05

Advanced Big Data Analysis & Science

Arcitura®
CERTIFIED
Big Data Scientist

OVERVIEW

This course delves into a range of advanced data analysis practices and analysis techniques that are explored within the context of Big Data. The course content focuses on topics that enable participants to develop a thorough understanding of statistical, modeling, and analysis techniques for data patterns, clusters and text analytics, as well as the identification of outliers and errors that affect the significance and accuracy of predictions made on Big Data datasets.

The following primary topics are covered:

- Modeling, Model Evaluation, Model Fitting and Model Overfitting
- Statistical Models, Model Evaluation Measures
- Cross-Validation, Bias-Variance, Confusion Matrix and F-Score
- Machine Learning Algorithms and Pattern Identification
- Association Rules and Apriori Algorithm
- Data Reduction, Dimensionality Feature Selection
- Feature Extraction, Data Discretization (Binning and Clustering)
- Advanced Statistical Techniques
- Parametric vs. Non-Parametric, Clustering vs. Non-Clustering
- Distance-Based, Supervised vs. Semi-Supervised
- Linear Regression and Logistic Regression for Big Data
- Classification Rules for Big Data
- Logistics Regression, Naïve Bayes, Laplace Smoothing, etc.
- Decision Trees for Big Data
- Tree Pruning, Feature Splitting, One Rule (1R) Algorithm
- Pattern Identification, Association Rules, Apriori Algorithm
- Time Series Analysis, Trend, Seasonality
- K Nearest Neighbor (kNN), K-means
- Text Analytics for Big Data
- Bag of Words, Term Frequency, Inverse Document Frequency, Cosine Distance, etc.
- Outlier Detection for Big Data
- Statistical, Distance-Based, Supervised and Semi-Supervised Techniques

Duration: 1 Day

TRAINING + CERTIFICATION

This course is part of a training and accreditation program through which official certification can be achieved and for which official training and certification badges can be issued.

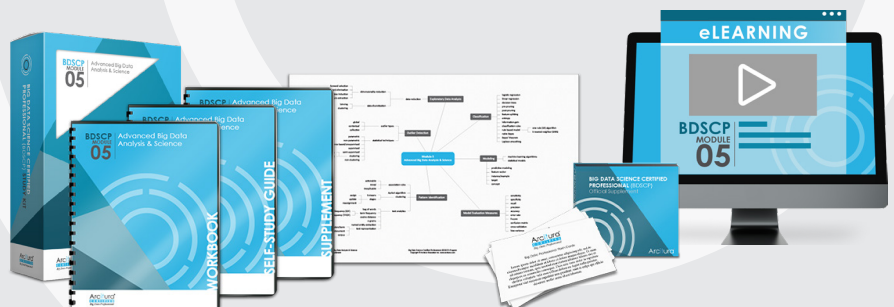


www.arcitura.com/bdscp/scientist

eLEARNING + PRINTED STUDY KITS

This Arcitura study kit includes the contents listed below. These course materials are available in full-color printed format, as well as in eLearning subscription format, via online access and offline file download.

- Workbook
- Self-Study Guide
- Supplement
- Mind Map Poster
- Flashcards
- Audio Tutor Recordings (usb)



Arcitura®

Pearson | VUE

Arcitura OnSite
EXAM PROCTORING

CClaim

Credly

BDSCP MODULE 05

Advanced Big Data Analysis & Science

	Certified Big Data Professional*	Certified Big Data Science Professional	Certified Big Data Scientist	Certified Big Data Consultant	Certified Big Data Engineer	Certified Big Data Architect	Certified Big Data Governance Specialist
MODULE 01 Fundamental Big Data	●	●	●	●	●	●	●
MODULE 02 Big Data Analysis & Technology Concepts	●	●	●	●	●	●	●
MODULE 03 Big Data Analysis & Technology Lab		●		●			
MODULE 04 Fundamental Big Data Analysis & Science			●	●			
MODULE 05 Advanced Big Data Analysis & Science			●				
MODULE 06 Big Data Analysis & Science Lab			●				
MODULE 07 Fundamental Big Data Engineering				●	●		
MODULE 08 Advanced Big Data Engineering					●		
MODULE 09 Big Data Engineering Lab					●		
MODULE 10 Fundamental Big Data Architecture						●	
MODULE 11 Advanced Big Data Architecture						●	
MODULE 12 Big Data Architecture Lab						●	
MODULE 13 Fundamental Big Data Governance							●
MODULE 14 Advanced Big Data Governance							●
MODULE 15 Big Data Governance Lab							●

* The Certified Big Data Professional designation is automatically issued when achieving any other BDSCP certification. It can also be achieved by receiving passing grades on Exams B90.01 + B90.02.