Certified Big Data Scientist

OEM: Arcitura • Duration: 3 Days (24 hrs) • Code: B90.BDS

COURSE MODULES & TOPICS

Module 1: Fundamental Big Data Science & Analytics

- Understanding Big Data
- Fundamental Big Data Terminology and Concepts
- Big Data Business Drivers and Technology Drivers
- Traditional Enterprise Technologies Related to Big Data
- OLTP, OLAP, ETL and Data Warehouses in relation to Big Data
- Characteristics of Data in Big Data Environments
- Dataset Types in Big Data Environments
- Structured, Unstructured and Semi-Structured Data
- Metadata and Data Veracity
- Fundamental Analysis and Analytics
- Quantitative and Qualitative Analysis
- Machine Learning Types
- Descriptive and Diagnostic Analytics
- Predictive and Prescriptive Analytics
- Business Intelligence and Big Data
- Data Visualization and Big Data
- Big Data Adoption and Planning Considerations

Module 2: Big Data Analysis & Technology Concepts

- Big Data Analysis Lifecycle (from Business Case Evaluation to Data Analysis and Visualization)
- A/B Testing and Correlation
- Regression and Heat Maps
- Time Series Analysis
- Network Analysis and Spatial Data Analysis
- Classification and Clustering
- Filtering, including Collaborative Filtering and Content-based Filtering
- Sentiment Analysis and Text Analytics
- Clusters and Processing Batch and Transactional Workloads
- How Cloud Computing relates to Big Data
- Foundational Big Data Technology Mechanisms

- Big Data Storage Devices and Processing Engines
- Resource Managers, Data Transfer Engines and Query Engines
- Analytics Engines, Workflow Engines and Coordinate Engines

Module 4: Big Data Analysis & Science

- Data Science, Data Mining & Data Modeling
- Big Data Dataset Categories
- High-Volume, High-Velocity, High-Variety, High-Veracity, High-Value Datasets
- Exploratory Data Analysis (EDA)
- EDA Numerical Summaries, Rules and Data Reduction
- EDA analysis types, including Univariate, Bivariate and Multivariate
- Essential Statistics, including Variable Categories and Relevant Mathematics
- Statistics Analysis, including Descriptive, Inferential, Covariance, Hypothesis Testing, etc.
- Measures of Variation or Dispersion, Interquartile Range & Outliers, Z-Score, etc.
- Probability, Frequency, Statistical Estimators, Confidence Interval, etc.
- Data Munging and Machine Learning
- Variables and Basic Mathematical Notations
- Statistical Measures and Statistical Inference
- Confirmatory Data Analysis (CDA)
- CDA Hypothesis Testing, Null Hypothesis, Alternative Hypothesis, Statistical Significance, etc.
- Distributions and Data Processing Techniques
- Data Discretization, Binning and Clustering
- Visualization Techniques, including Bar Graph, Line Graph, Histogram, Frequency Polygons, etc.
- Prediction Linear Regression, Mean Squared Error and Coefficient of Determination R2, etc.
- Clustering k-means, Cluster Distortion, Missing Feature Values, etc.
- Numerical Summaries

Module 5: Advanced Big Data Analysis & Science

- 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

Module 6: Big Data Analysis & Science Lab

- Reading Exercise 6.1: TMC Case Study Background
- Lab Exercise 6.2: Analysis for Enhancing Product Quality
- Lab Exercise 6.3: Analysis for Lowering Total Cost of Ownership
- Reading Exercise 6.4: PLGM Case Study Background
- Lab Exercise 6.5: Analysis for High-Yield Marketing Plan
- Lab Exercise 6.6: Analyze Items Layout and Credit Card Data
- Reading Exercise 6.7: LHL Case Study Background
- Lab Exercise 6.8: Enhance Patient Diagnosis Capability
- Reading Exercise 6.9: SWP Case Study Background
- Lab Exercise 6.10: Enhance Risk Management and Understand Demand Patterns
- is authored by a dedicated courseware development team
- has a self-test, accreditation exam and professional certification
- is available via two different eLearning platforms
- undergo a common development process
- are authored to be consistent in quality, structure and style
- share a common vocabulary and symbol notation
- are authored in collaboration with subject matter experts
- About Arcitura
- Instructor-Led Training & Coaching
- eLearning with Arcitura
- Course & Certification Tracks
- Exams & Proctoring
- Digital Accreditations
- Trainer Development
- Partner Program
- Partner Portal
- Privacy Policy
- Candidate Agreement
- Logo Guidelines
- Contact
- Help
- Arcitura on LinkedIn