

Certified Big Data Consultant

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

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 3: Big Data Analysis & Technology Lab

- Reading Exercise 3.1: Case Study Background PLGM
- Lab Exercise 3.2: Plan the Big Data BI Environment
- Lab Exercise 3.3: Analyze Customer Loyalty Data
- Lab Exercise 3.4: Alleviate Customer Dissatisfaction
- Lab Exercise 3.5: Improve PLGM's On-Line Sales
- Reading Exercise 3.6: Case Study Background LHL
- Lab Exercise 3.7: Plan the Data Integration and Reporting Environment
- Lab Exercise 3.8: Develop a Treatment Personalization Capability
- Lab Exercise 3.9: Enhance LHL's Research Capability
- Reading Exercise 3.10: Case Study Background SWP
- Lab Exercise 3.11: Smart Meter Data Analysis
- Lab Exercise 3.12: Enhance Electricity Demand Prediction Capability
- Lab Exercise 3.13: Asset Management and Risk Identification Capability

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 R^2 , etc.
- Clustering k-means, Cluster Distortion, Missing Feature Values, etc.
- Numerical Summaries

Module 11: Fundamental Big Data Engineering

- Big Data Engineering Techniques and Challenges
- Big Data Storage, including Sharding, Replication, CAP Theorem, ACID and BASE
- Master-Slave, Peer-to-Peer Replication, Combining Replication with Sharding
- Big Data Storage Requirements, Scalability, Redundancy and Availability
- Fast Access, Long-term Storage, Schema-less Storage and Inexpensive Storage
- On-Disk Storage, including Distributed File System and Databases
- Introduction to NoSQL and NewSQL
- NoSQL Rationale and Characteristics
- NoSQL Database Types, including Key-Value, Document, Column-Family and Graph Databases
- Big Data Processing Engines
- Distributed/Parallel Data Processing, Schema-less Data Processing
- Multi-Workload Support, Linear Scalability and Fault-Tolerance
- Big Data Processing Requirements, including Batch, Cluster and Realtime Modes
- MapReduce for Big Data Processing, including Map, Combine, Partition, Shuffle and Sort and Reduce
- MapReduce Algorithm Design
- Task Parallelism, Data Parallelism
- 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