Big Data Analysis & Advanced Data Science

Training & Certification Guide

Arcitura® TRAINING — Advanced Data Science

Arcitura® TRAINING — Advanced Big Data Analysis

Arcitura® CERTIFIED — Big Data Scientist

itcc

PCC PROFESSIONAL CERTIFICATION COALITION

Pearson | VUE

OnVUE

Acclaim

Credly by Pearson

Arcitura® DIGITAL

About the Next-Gen Data Science Academy

The Next-Gen Data Science Academy from Arcitura provides formal education and accreditation programs dedicated to the fields of Artificial Intelligence, Machine Learning, Big Data and general Data Science, including analytics and analysis, architecture, engineering and governance.

The Next-Gen Data Science Academy curriculum is comprised of 24 course modules and 9 certification tracks. Exams are available worldwide via Pearson VUE testing centers, as well as via Pearson VUE OnVue online proctoring and on-site delivery by Certified Trainers. Achieving a passing grade on the required exam(s) achieves a certification for which a digital accreditation certificate is automatically issued by Arcitura and a digital certification badge is issued by Acclaim/Credly.

TABLE OF CONTENTS

HOW TO TAKE THIS COURSE

ARCITURA eLEARNING OPTIONS

To give you the most flexibility to achieve your learning goals and accommodate your preferences, this course is made available via two Arcitura eLearning solutions: An interactive environment with graded exercises and a graded self-test, as well as a study kit account that supports online/offline access and custom annotations.

To learn more, visit: www.arcitura.com/elearning

To enroll, visit: digital.arcitura.com/courses

ONSITE WORKSHOPS

Private onsite workshops can be delivered by Certified Trainers from Arcitura and authorized training partners for both small and large groups. Each workshop participant can receive access to the full Arcitura Digital eLearning course materials. Each workshop agenda, format and schedule can be tailored to client requirements.

To learn more, visit: www.arcitura.com/workshops

VIRTUAL WORKSHOPS

Private virtual workshops can be delivered by Certified Trainers for small and large groups, as well as individual participants. Workshop participants receive access to the course materials via the Arcitura Digital eLearning platform. Virtual workshop agendas can be tailored with greater flexibility to accommodate more distributed and fragmented training schedules.

To learn more, visit: www.arcitura.com/workshops

Several additional learning and exam preparation products and services are available, including coaching, exam prep kits and digital downloads. See the Training and Exam Preparation Resources page for details.

Arcitura® DIGITAL
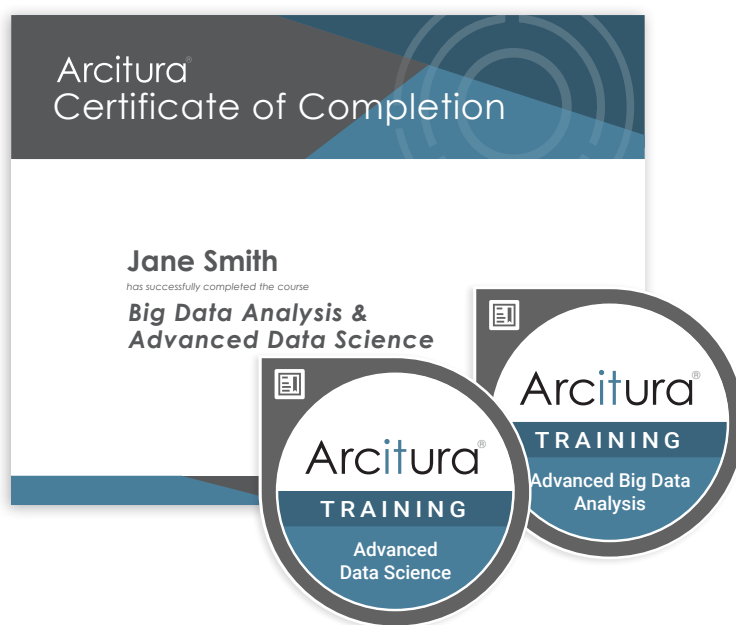
HOW TO GET STARTED

Welcome to the Big Data Analysis & Advanced Data Science course! This course is comprised of a set of modules. Each module has a set of lessons and is further supplemented with exercises to help reinforce your understanding of key topics. Upon completing the course, you can optionally proceed to prepare yourself for the certification exam (as explained on the *How to get Certified* page).

Additional resources are available to assist you with completing this course, including downloadable digital course files, printed course materials, coaching hours and instructor-led training services (as explained on the *Training and Exam Preparation Resources* page.)
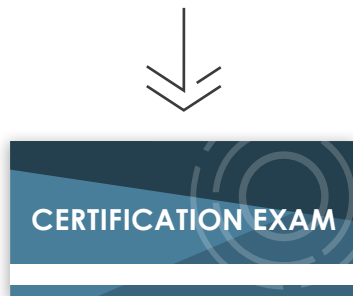
PREREQUISITE

The prerequisite for this course is the completion of the Big Data Analytics & Fundamental Data Science course. It is recommended that you complete the Big Data Analytics & Fundamental Data Science course prior to starting the Big Data Analysis & Advanced Data Science course.

Arcitura CERTIFIED
Big Data Scientist

A Certified Big Data Scientist has knowledge of a range of analysis and analytics techniques, as well as the processes required for processing large volumes of complex data to drive decision-making.

Module 04

Module 05

Module 06

CERTIFICATION EXAM

Arcitura®
Certified Big Data Scientist

Arcitura®
Certified Big Data Scientist

Jane Smith
has successfully completed the requirements to be recognized as a certified Big Data Scientist.

HONORS

Arcitura CERTIFIED
Big Data Scientist

www.arcitura.com

Arcitura®
CERTIFIED
Big Data Scientist

HOW TO GET CERTIFIED

You can become a Certified Big Data Scientist! This course can prepare you for the official Big Data Scientist Certification exam, which can be taken worldwide at Pearson VUE testing centers, via Pearson VUE online proctoring and/or Arcitura direct proctoring.

Upon attaining a passing grade on the certification exam (and fulfilling any other prerequisite exam requirements) you will achieve the Big Data Scientist Certification, after which you will automatically receive an official digital Accreditation Certificate and a digital Certification Badge from Acclaim/Credly with an account that supports the online verification of your certification status. Digital accreditation certificates and badges are free of charge.

Additional resources are available to assist you with preparing for the certification exam, including practice exam questions, downloadable digital course files, printed course materials, coaching hours and instructor-led training services (as explained on the *Training and Exam Preparation Resources* page.)

Pearson | VUE   OnVUE   Acclaim   Credly by Pearson

Arcitura®
TRAINING
Advanced Data Science

The Big Data Analysis & Advanced Data Science course provides comprehensive coverage of contemporary Big Data analysis and analytics practices and advanced data science techniques and processes.

COURSE MODULE OUTLINES

The Big Data Analysis & Advanced Data Science course is comprised of the following course modules. Outlines for these course modules are provided on the subsequent pages.
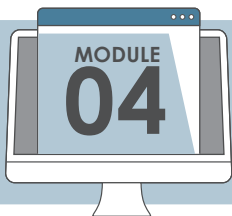
MODULE 04

Big Data Analysis & Science
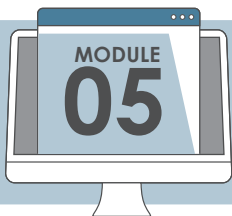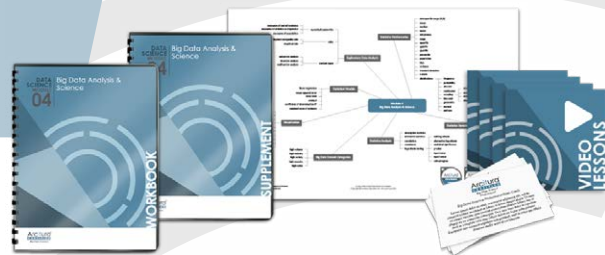
MODULE 05

Advanced Big Data Analysis & Science

MODULE 06

Big Data Analysis & Science Lab

Big Data Analysis & Science

MODULE 04

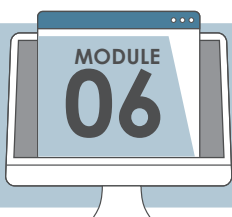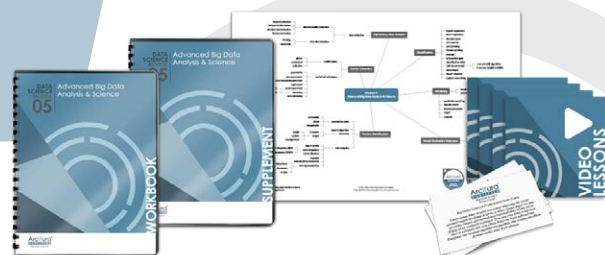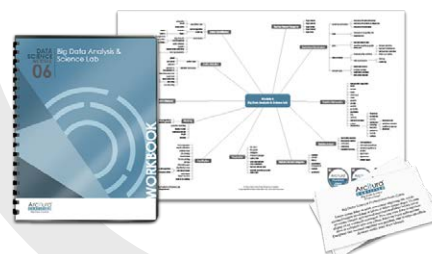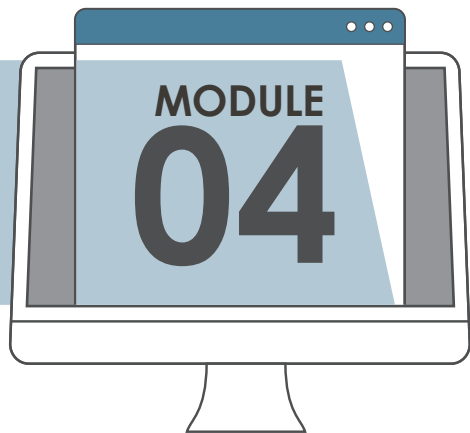This course module provides an in-depth overview of essential topic areas pertaining to data science and analysis techniques relevant and unique to big data with an emphasis on how analysis and analytics need to be carried out individually and collectively in support of the distinct characteristics, requirements and challenges associated with big data datasets.

The following primary topics are covered:

- 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

Advanced Big Data Analysis & Science
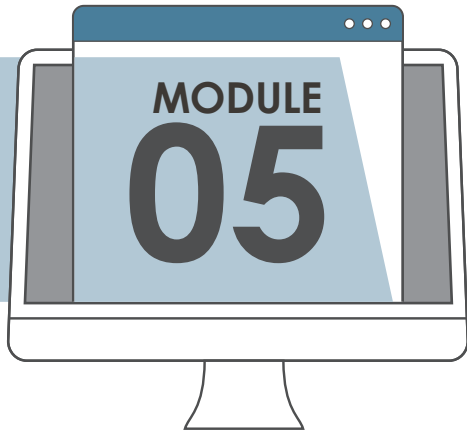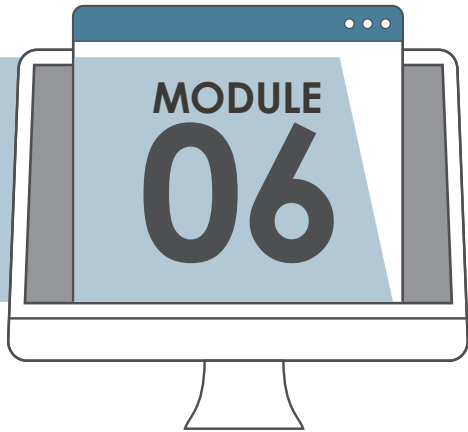
MODULE 05

This course module delves into a range of advanced data analysis practices and analysis techniques that are explored within the context of big data. The module 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

Big Data Analysis & Science Lab

MODULE 06

////////////

This course module covers a series of exercises and problems designed to test the participant's ability to apply knowledge of topics covered previously in module modules 4 and 5. Completing this lab will help highlight areas that require further attention, and will further prove proficiency in big data analysis and science practices as they are applied and combined to solve real-world problems.

The following exercises are provided:

- 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

TRAINING AND EXAM PREPARATION RESOURCES

You can supplement this course with a number of available resources to assist with both learning and exam preparation. Contact info@arcitura.com with any questions.

Certification Exam Prep Kit

A set of additional practice questions is available to support exam preparation.
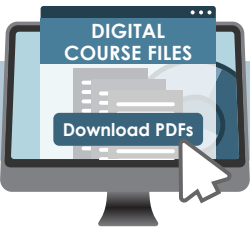
Digital Course Files

For each course you can order a set of downloadable digital course materials comprised of printable, watermarked workbook and poster PDF files.

Printed Course Materials

The printed workbooks and posters for each course can be ordered in B&W and full-color, and can be shipped worldwide.

One-on-One Coaching

Certified Trainers are available to provide online coaching on an hourly basis and in all time zones.

Instructor-Led Training

Certified Trainers are available to provide virtual and onsite training workshops for this and other Arcitura courses.

DIGITAL TRANSFORMATION PROFESSIONAL ACADEMY

Arcitura® DIGITAL

COURSES → CERTIFICATIONS ↓	Digital Transformation / Digital Transformation Specialist	Digital Transformation: Fundamental Technology / Digital Transformation Technology Professional	Digital Transformation: Advanced Technology & Architecture / Digital Transformation Technology Architect	Digital Transformation: Fundamental Data Science / Digital Transformation Data Science Professional	Digital Transformation: Advanced Data Science / Digital Transformation Data Scientist	Digital Transformation: Fundamental Security / Digital Transformation Security Professional	Digital Transformation: Advanced Security / Digital Transformation Security Specialist	Digital Transformation: Fundamental Intelligent Automation / Digital Transformation Intelligent Automation Professional	Digital Transformation: Advanced Intelligent Automation / Digital Transformation Intelligent Automation Specialist
MODULE 01 — Fundamental Digital Transformation	●	○	○	○	○	○	○	○	○
MODULE 02 — Digital Transformation in Practice	●	○	○	○	○	○	○	○	○
MODULE 03 — Fundamental Cloud Computing		●	○						
MODULE 04 — Fundamental Blockchain		●	○			●	○		
MODULE 05 — Fundamental IoT		●	○						
MODULE 06 — Cloud Architecture			●						
MODULE 07 — Blockchain Architecture			●				●		
MODULE 08 — IoT Architecture			●						
MODULE 09 — Fundamental Big Data Analysis & Analytics				●	○				
MODULE 10 — Fundamental Machine Learning				●	○				
MODULE 11 — Fundamental AI				●	○				
MODULE 12 — Advanced Big Data Analysis & Analytics					●				
MODULE 13 — Advanced Machine Learning					●				
MODULE 14 — Advanced AI					●				
MODULE 15 — Fundamental Cybersecurity						●	○		
MODULE 16 — Advanced Cybersecurity							●		
MODULE 17 — Fundamental RPA								●	○
MODULE 18 — Advanced RPA & Intelligent Automation									●
MODULE 19 — Fundamental AI Decisioning								●	○
MODULE 20 — Advanced AI Decisioning									●

Gray circles indicate prerequisite modules.

Pearson | VUE   OnVUE   Acclaim   Credly by Pearson

NEXT-GEN IT ACADEMY

COURSES	DevOps	Blockchain Architecture	IoT Architecture	Cybersecurity	Robotic Process Automation	Digital Business Technology	Containerization Architecture	Quantum Computing
CERTIFICATIONS	DevOps Specialist	Blockchain Architect	IoT Architect	Cybersecurity Specialist	RPA Specialist	Digital Business Technology Professional	Containerization Architect	Quantum Computing Specialist
DevOps MODULE 01 Fundamental DevOps	●							
MODULE 02 DevOps in Practice	●							
MODULE 03 DevOps Lab	●							
Blockchain MODULE 01 Fundamental Blockchain		●						
MODULE 02 Blockchain Technology & Architecture		●						
MODULE 03 Blockchain Technology & Architecture Lab		●						
Internet of Things MODULE 01 Fundamental IoT			●					
MODULE 02 IoT Technology & Architecture			●					
MODULE 03 IoT Technology & Architecture Lab			●					
Cybersecurity MODULE 01 Fundamental Cybersecurity				●				
MODULE 02 Advanced Cybersecurity				●				
MODULE 03 Cybersecurity Lab				●				
Robotic Process Automation MODULE 01 Fundamental RPA					●			
MODULE 02 Advanced RPA & Intelligent Automation					●			
MODULE 03 RPA Lab					●			
Digital Business Technology MODULE 01 Business Automation Technology Overview						●		
MODULE 02 Data Science Technology Overview						●		
MODULE 03 Digital & Security Technology Overview						●		
Containerization MODULE 01 Fundamental Containerization							●	
MODULE 02 Containerization Technology & Architecture							●	
MODULE 03 Containerization Technology & Architecture Lab							●	
Quantum Computing MODULE 01 Fundamental Quantum Computing								●
MODULE 02 Advanced Quantum Computing								●
MODULE 03 Quantum Computing Lab								●

Next-Gen Data Science Academy

COURSES / CERTIFICATIONS	Big Data Analytics & Fundamental Data Science / Big Data Science Professional	Big Data Analysis & Advanced Data Science / Big Data Scientist	Data Science Professional Consulting / Data Science Consultant	Machine Learning / Machine Learning Specialist	Artificial Intelligence / Artificial Intelligence Specialist	Big Data Engineering / Big Data Engineer	Big Data Architecture / Big Data Architect	Data Science Governance / Data Science Governance Specialist	AI Decisioning / AI Decisioning Specialist
MODULE 01 — Fundamental Big Data Science & Analytics	●	○	○			○	○	○	○
MODULE 02 — Big Data Analysis & Technology Concepts	●	○	○			○	○	○	○
MODULE 03 — Big Data Analysis & Technology Lab	●		○						
MODULE 04 — Big Data Analysis & Science		●							
MODULE 05 — Advanced Big Data Analysis & Science		●							
MODULE 06 — Big Data Analysis & Science Lab		●							
MODULE 07 — Fundamental Machine Learning			●	●					
MODULE 08 — Advanced Machine Learning				●					
MODULE 09 — Machine Learning Lab				●					
MODULE 10 — Fundamental Artificial Intelligence			●		●				
MODULE 11 — Advanced Artificial Intelligence					●				
MODULE 12 — Artificial Intelligence Lab					●				
MODULE 13 — Fundamental Big Data Engineering						●			
MODULE 14 — Advanced Big Data Engineering						●			
MODULE 15 — Big Data Engineering Lab						●			
MODULE 16 — Fundamental Big Data Architecture							●		
MODULE 17 — Advanced Big Data Architecture							●		
MODULE 18 — Big Data Architecture Lab							●		
MODULE 19 — Fundamental Data Science Governance for Big Data, Machine Learning & AI								●	
MODULE 20 — Advanced Data Science Governance for Big Data, Machine Learning & AI								●	
MODULE 21 — Data Science Governance Lab for Big Data, Machine Learning & AI								●	
MODULE 22 — Fundamental AI Decisioning									●
MODULE 23 — Advanced AI Decisioning									●
MODULE 24 — AI Decisioning Lab									●

Gray circles indicate prerequisite modules.

Pearson | VUE   onVUE   Acclaim   Credly by Pearson

Cloud School

COURSES / CERTIFICATIONS	Cloud Computing / Cloud Technology Professional	Cloud Computing Professional Consulting / Cloud Computing Consultant	Cloud Architecture / Cloud Architect	Cloud Security / Cloud Security Specialist	Cloud Governance / Cloud Governance Specialist	Cloud Storage / Cloud Storage Specialist	Cloud Virtualization / Cloud Virtualization Specialist
MODULE 01 Fundamental Cloud Computing	●	○	○	○	○	○	○
MODULE 02 Cloud Technology Concepts	●	○	○	○	○	○	○
MODULE 03 Cloud Technology Lab	●	○					
MODULE 04 Fundamental Cloud Architecture		●	●				
MODULE 05 Advanced Cloud Architecture			●				
MODULE 06 Cloud Architecture Lab			●				
MODULE 07 Fundamental Cloud Security		●		●			
MODULE 08 Advanced Cloud Security				●			
MODULE 09 Cloud Security Lab				●			
MODULE 10 Fundamental Cloud Governance					●		
MODULE 11 Advanced Cloud Governance					●		
MODULE 12 Cloud Governance Lab					●		
MODULE 13 Fundamental Cloud Storage						●	
MODULE 14 Advanced Cloud Storage						●	
MODULE 15 Cloud Storage Lab						●	
MODULE 16 Fundamental Cloud Virtualization							●
MODULE 17 Advanced Cloud Virtualization							●
MODULE 18 Cloud Virtualization Lab							●

Gray circles indicate prerequisite modules.

SERVICE TECHNOLOGY SCHOOL

Arcitura DIGITAL

COURSES →	Fundamental Microservices & Service Technology	Fundamental SOA Design with Services & Microservices	SOA Analysis & Modeling with Services & Microservices	SOA Design & Architecture with Services & Microservices	Microservice Design & Architecture	Microservice Professional Consulting	Service API Design & Management	Service Governance & Project Delivery	Security for Microservices & SOA
CERTIFICATIONS	Microservice Professional	SOA Professional	SOA Analyst	SOA Architect	Microservice Architect	Microservice Consultant	Service API Specialist	Service Governance Specialist	Service Security Specialist
MODULE 01 — Fundamental SOA, Services & Microservices	●	●	○	○	○	○	○	○	○
MODULE 02 — Microservice Technology Concepts	●			○	○	○	○		○
MODULE 03 — Design & Architecture with SOA, Services & Microservices		●	○	●				○	
MODULE 04 — Fundamental SOA Analysis & Modeling with Services & Microservices			●						
MODULE 05 — Advanced SOA Analysis & Modeling with Services & Microservices			●						
MODULE 06 — SOA Analysis & Modeling Lab with Services & Microservices			●						
MODULE 07 — Advanced SOA Design & Architecture with Services & Microservices				●					
MODULE 08 — SOA Design & Architecture Lab with Services & Microservices				●					
MODULE 09 — Fundamental Microservice Architecture & Containerization					●	●			
MODULE 10 — Advanced Microservice Architecture & Containerization					●				
MODULE 11 — Microservice Architecture & Containerization Lab					●				
MODULE 12 — Fundamental Service API Design & Management						●	●		
MODULE 13 — Advanced Service API Design & Management							●		
MODULE 14 — Service API Design & Management Lab							●		
MODULE 15 — Fundamental Service Governance & Project Delivery								●	
MODULE 16 — Advanced Service Governance & Project Delivery								●	
MODULE 17 — Service Governance & Project Delivery Lab								●	
MODULE 18 — Fundamental Security for Services, Microservices & SOA						●			●
MODULE 19 — Advanced Security for Services, Microservices & SOA									●
MODULE 20 — Security Lab for Services, Microservices & SOA									●

Gray circles indicate prerequisite modules.

Arcitura®