Big Data Scientist

Durée: 4 Jours   Réf de cours: BDS

Résumé:

This 4-day Big Data Scientist course is a continuation of the Big Data Fundamentals (BDF) course and consists of 4 modules. After each module the participant can take the corresponding exam. If all exams are passed, the participant will be Certified Big Data Scientist. The course is scheduled in 4 blocks of 1 day, spread over approximately 4 weeks.
The modules are part of the Big Data Science Certified Professional (BDSCP) curriculum of Arcitura Education. The Big Data Science Certified Professional (BDSCP) program from Arcitura is dedicated to excellence in the fields of Big Data science, analysis, analytics, business intelligence, and technology architecture, as well as design, development, and governance.

Public visé:

This course is intended for anyone who, after following the Big Data Fundamentals course, has come to the conclusion that he or she wishes or needs to gain more insight into the field of Big Data.
In this course, the emphasis is more on broadening than deepening knowledge; the participants are not trained as specialists but as generalists. Gaining an overview is important, because anyone who lacks one cannot successfully become part of a Big Data team.

Pré-requis:

BDF, Big Data Fundamentals

Test et certification

This course consists of 4 modules. Each module can be concluded with an exam. The exams are optional and are not included in the price.

If you want to take the 4 exams, you need to indicate this in advance.

It concerns the following 4 exams (via Pearson Vue) of Arcitura Education:

- B90.02, Big Data Analysis & Technology Concepts
- B90.04, Fundamentals Big Data Analysis & Science
- B90.05, Advanced Big Data Analysis & Science
- B90.06, Big Data Analysis & Science Lab

Contenu:

Module 1: Big Data Analysis ; Technology Concepts

- Big Data Analysis Lifecycle (from business case evaluation to data analysis and visualization)
- A/B Testing, Correlation
- Regression, Heat Maps
- Time Series Analysis
- Network Analysis
- Spatial Data Analysis
- Classification, Clustering
- Outlier Detection
- Filtering (including collaborative filtering ; content-based filtering)
- Natural Language Processing
- Sentiment Analysis, Text Analytics
- File Systems ; Distributed File Systems, NoSQL
- Distributed ; Parallel Data Processing,
- Processing Workloads, Clusters
- Cloud Computing ; Big Data
- Foundational Big Data Technology Mechanisms

Module 2: Fundamentals Big Data Analysis ; Science

- Data Science, Data Mining ; Data Modeling
- Big Data Dataset Categories
- Exploratory Data Analysis (EDA) (including numerical summaries, rules ; data reduction)
- EDA analysis types (including univariate, bivariate ; multivariate)
- Essential Statistics (including variable categories ; relevant mathematics)
- Statistics Analysis (including descriptive, inferential, correlation, covariance ; hypothesis testing)
- Data Munging ; Machine Learning
- Variables ; Basic Mathematical Notations
- Statistical Measures ; Statistical Inference
- Distributions ; Data Processing Techniques
- Data Discretization, Binning, Clustering
- Visualization Techniques ; Numerical Summaries
- Correlation for Big Data
- Time Series Analysis for Big Data

Module 3: Advanced Big Data Analysis ; Science

- Statistical Models, Model Evaluation Measures (including cross-validation, bias-variance, confusion matrix ; f-score)
- Machine Learning Algorithms, Pattern Identification (including association rules ; apriori algorithm)
- Advanced Statistical Techniques (including parametric vs. non-parametric, clustering vs. non-clustering distance-based, supervised vs. semi-supervised)
- Linear Regression ; Logistic Regression for Big Data
- Decision Trees for Big Data
- Classification Rules for Big Data
- K Nearest Neighbor (kNN) for Big Data
- Naïve Bayes for Big Data
- Association Rules for Big Data
- K-means for Big Data
- Text Analytics for Big Data
- Outlier Detection for Big Data

Module 4: Big Data Analysis ; Science lab

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 course modules 4 and 5. Completing this lab will help highlight areas that require further attention, and will further prove hands-on proficiency in Big Data analysis and science practices as they are applied and combined to solve real-world problems.

As a hands-on lab, this course incorporates a set of detailed exercises that require participants to solve various inter-related problems, with the goal of fostering a comprehensive understanding of how different data analysis techniques can be applied to solve problems in Big Data environments and used to make significant, relevant predictions that offer increased business value.

Méthodes pédagogiques :

The course is mainly theoretical, abstract and vendor-neutral. Apart from Excel, no specific software is used. However, it is desirable to bring your Windows laptop for the exercises.

Autres moyens pédagogiques et de suivi:

• Compétence du formateur : Les experts qui animent la formation sont des spécialistes des matières abordées et ont au minimum cinq ans d'expérience d'animation. Nos équipes ont validé à la fois leurs connaissances techniques (certifications le cas échéant) ainsi que leur compétence pédagogique.
• Suivi d'exécution : Une feuille d'émargement par demi-journée de présence est signée par tous les participants et le formateur.
• Modalités d'évaluation : le participant est invité à s'auto-évaluer par rapport aux objectifs énoncés.
• Chaque participant, à l'issue de la formation, répond à un questionnaire de satisfaction qui est ensuite étudié par nos équipes pédagogiques