



# Big Data & Data Science Program Diploma Courses

Date: March, 2019 – v 3.0

## Diploma Structure

The Big Data & Data Science Diploma requires the attendance of 4 courses and 1 hands-on group project according to the following structure:

### Semester #1 (2 Courses)

- 1- Introduction to Big Data, Developing with Spark and Hadoop (42 Hours, 14 Lectures)
- 2- Introduction to Machine Learning and Statistical Analysis (42 Hours, 14 Lectures)

### Semester #2 (2 Courses)

- 3- Advanced Big Data Analytics Technologies and Applications (42 Hours, 14 Lectures)
- 4- Only 1 of the 3 following courses:
  - Practical Data Mining (42 Hours, 14 Lecture)

OR

- Practical Data Science Using Machine Learning Technique (42 Hours, 14 Lectures)

OR

- Selected topics in Deep Learning (42 Hours, 14 Lectures)

### Semester #3 (Final Project)

- Hands-on group project based on real life use case (14 Weeks of Mentoring)

Please refer to **Appendix A** for the description of each of those courses.

## Important Notes

- All enrollments are subject to the admission rules and acceptance criteria of Nile University and the Big Data and Data Science Program.
- The default training location in Nile University premises and any change will be decided upon case by case by the program management team.
- Timing, lecture distribution, assigned instructors and schedules will be assigned and announced to students upon registration completion subject to Nile University and the program administrative decisions.
- The courses details and outlines might get changed due to continuous development and enhancements to cope with trending theories, technologies, methods and applications in this domain.

For more details and pricing, please contact us: [bigdata@nu.edu.eg](mailto:bigdata@nu.edu.eg)

## Appendix A: Course Descriptions

### CIT-652: Advanced Big Data Analytics Technologies and Applications (42 Hours, 14 Lecture)

#### Description

This is an applied course where students can develop on their combined knowledge of Big Data technologies (e.g. Hadoop, Spark, etc.) and Data Science (e.g. Statistics, Machine Learning, etc.) and understand how such combination is used to solve real-world applications. In addition to this main goal, the course has the additional goal of familiarizing students with the latest technological and scientific trends in the field and how Big Data and data science are used in modern business enterprises. Use cases of real problems such as networking traffic, text analytics, and financial applications will be addressed in this course.

#### Pre-requisites

- CIT-650: Introduction to Big Data, Hadoop and Spark  
(Or equivalent knowledge subject to NU evaluation)

#### Reference Textbook

Ryza, Sandy, Uri Laserson, Sean Owen and Josh Wills. Advanced Analytics with Spark Patterns for Learning from Data at Scale. O'Reilly Media, April 2015.

#### Course Outlines

- Analyzing Data with Scala and Spark
- Predicting Forest Cover with Decision Trees
- Anomaly Detection in Network Traffic with K-means Clustering
- Understanding Wikipedia with Latent Semantic Analysis
- Analyzing Co-occurrence Networks with GraphX
- Geospatial and Temporal Data Analysis on Taxi Trip Data
- Estimating Financial Risk through Monte Carlo Simulation