



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

Appendix A: Course Descriptions

CIT-690B: Practical Data Science Using Machine Learning Techniques (42 Hours, 14 Lectures)

Description

Data Sciences is a fast evolving practice that apply several sciences, theories and techniques in solving complex data-related problems and develop applications that support transforming the way organizations do their activities. Machine learning techniques are at the core of data sciences and have proven great value in the context of the practical applications of Big Data solutions. In this course, the students will be able to link the machine learning theories and methods in a practical real-life use case context. The hands-on labs will reinforce the concepts learned from the Introduction to Machine Learning course (CIT-651) with deeper focus in applying them to enable customers realize business value.

Furthermore, students will learn how the actual customer engagements in this filed works including consulting and implementation approaches. The course will demonstrate tools and technologies – both open source and commercial – like R, Java/Spring, Weka, Hadoop, Spark, Giraph, Cloud Foundry, Madlib and Greenplum applied to practical situations.

Pre-requisites

- CIT-651: Introduction to Machine Learning and Statistical Analysis
(Or equivalent knowledge subject to NU evaluation)

Reference Textbooks

Main reference: “Learning From Data”, Yaser S. Abu-Mostafa, Malik Magdon-Ismael, Hsuan-Tien Lin, March 27, 2012.

Other References including:

- Articles, Blogs, Forums...
- Technical Papers, Research Papers, Text Books...
- Industry Analysts, Vendors, Experts...
- Tutorials, PoV, Benchmarks, Docs, Videos...

Course Outline

Module 1: Big Data and Data Science: Use Cases

- Walk through example use cases

- The use case: solution overview and architecture
- The practice of the Data Sciences vs Traditional DW/BI
- Data Sciences and Big Data Applications: Value to the Business

Module 2: Data Preprocessing

- Features Extraction and Transformations
- Dimensionality Reduction
- Visualization and Exploratory Data Analysis
- Data Integration, Quality and Implications.
- Handling Big Datasets

Module 3: Advanced Data Analysis Methods with Applications

- Unstructured Data Methods
- Association Rules: Understand Customers Behavior
- Clustering Techniques: Optimized Logistics
- Classifications Methods: Prediction of Traffic Status
- Network Analysis Techniques: Discover Social Patterns
- Big Graph: Analyzing Electric Power Grids
- Ensemble Learning techniques

Module 4: Practicing Big Data Sciences in Real Life

- The Data Sciences modeling lifecycle
- Machine Learning modeling for Big Data applications
- Data Sciences application implementation lifecycle
- Deep Learning for Complex Data Science Models
- Machine Learning Agile modeling approach
- Data Driven Transformation for Organizations
- Consulting Skills for the Data Sciences and Big Data Solutions
- Deployment Considerations for the Big Data Platforms