

Module 5

Data Analytics

Course Overview

Juan Carlos Trujillo – Lucentia (jtrujillo@dlsi.ua.es)
Alejandro Maté – Lucentia (amate@dlsi.ua.es)



Data Analytics can be divided into two main tasks:

- **Data preparation:**
 - Data sources selection, data extraction, transformation, loading, analytic schema creation, visualizations...
 - Essentially, all the necessary tasks for being able to analyze data
- **Data analysis:**
 - Statistics, econometric & mathematical models, machine learning...
 - Essentially, all the techniques related to analyzing the data

This module focuses on the **data preparation aspect** that is necessary before carrying out data analysis and the **initial analysis**

Course Outline

- Lesson 1: Introduction to Analytics: Multidimensional modeling
 - Conceptual multidimensional design
 - Logical multidimensional design
 - Multidimensional model implementation
- Lesson 2: ETL Processes
 - ETL processes fundamentals and design
 - ETL processes implementation
- Lesson 3: Building a Data Warehouse solution
 - ETL processes implementation
 - Implementing Data Warehouses
- Lesson 4 Analytics over Big Data sources: Data Science fundamentals:
 - Introduction to data science: roles & steps
 - Big Data sources
- Lesson 5: Big Data analytics
 - Data science tools usage
 - Data source profiling & analysis
- Lesson 6: Machine Learning for Data Analytics
 - Introduction to Machine Learning
 - Machine Learning Analytics

Schedule

18:00
19:00
20:00
21:00

6 Jan	7 Jan	8 Jan	9 Jan	10 Jan
Holiday	Holiday	MD Modeling	ETL Processes	Building a DW
		MD with ORACLE	ETL with PDI	ETL with PDI

18:00
19:00
20:00
21:00

13 Jan	14 Jan	15 Jan	16 Jan	17 Jan
Student's Work Rec. Lectures	Student's Work Rec. Lectures	Data Science Introduction & Data sources selection	Data Profiling & Analysis	Machine Learning for Data Analytics
			Data Processing Notebooks	

18:00
19:00
20:00
21:00

23 Jan
Student's Work

Student's Work

- Lesson 1: Introduction to Analytics: Multidimensional (MD) modeling
 - MD model design
 - MD implementation by using ORACLE
- Lesson 2: ETL Processes
 - ETL processes implementation by using Pentaho Data Integration (PDI)
- Lesson 3: Building a Data Warehouse solution
 - ETL processes implementation by using Pentaho Data Integration (PDI)
 - Implementing a Data Warehouse repository
- Lesson 4: Analytics over Big Data sources: Data Science fundamentals:
 - Big Data source search & selection
- Lesson 5: Big Data analytics
 - Big Data source profiling & analysis using Jupyter
- Lesson 6: Building a Machine Learning Model
 - Deployment of an ML model by using Jupiter notebook

Format and Deadlines:

- All tasks will be submitted as PDF files through the Virtual Campus
- Due date:

Two weeks after the class for the tasks

Module 5

Data Analytics

Course Overview

Juan Carlos Trujillo – Lucentia (jtrujillo@dlsi.ua.es)
Alejandro Maté – Lucentia (amate@dlsi.ua.es)

