

Big Data Analytics @ BSC

Monday 03 February 2020 - Friday 07 February 2020

Barcelona BSC Campus Nord

Agenda

Day 1 (Feb 3)

9:30 – 13:00 Introduction to Big Data (David Carrera, Data Centric Computing Group Manager, BSC)

The goal of this session is to introduce the students in the technologies associated with Big Data: data challenges, cloud computing, processing, and internet of things. An overview of the technologies will be provided, both from a technical and from a business model point of view.

11:00 - 11:30 Coffee break

13:00 – 14:00 Lunch Break

14:00 – 16:00 Practical Data Analytics for Solving Real World Problems (Carlos Carrasco, Data Pre&Post Processing, BSC)

Data analytics has changed the way we make decisions. We see the benefits and the advances in many fields that go from financial to medical and industrial applications due to the integration of advanced data analytics. In this course we will propose practical tips gained through our experience at BSC in big data analytics projects. We will also discover how to overcome some of the most challenging tasks in practical data analytics.

16:00 – 16:30 Coffee break

16:30 – 18:00 Hands-on

This session will focus on several key methods and algorithms (both serial and parallel) that enable to discover global properties on data while dealing with Big Data:

Network Science

Multi Constrained and Multi-Objective Optimization

Examples using the above approaches and some hands-on exercise

Day 2 (Feb 4)

9:30 – 13:00 Big Data Management

11:00 - 11:30 Coffee break

Hands-on exercise

13:00 – 14:00 Lunch Break

14:00 - 16:00 NoSQL databases (Oscar Romero, Dept. of Service and Information System Engineering, UPC-BarcelonaTech)

The relational model has dominated data storage systems since the mid 1970s. However, the changing storage needs over the past decade have given rise to new models for storing data, collectively known as NoSQL. In this presentation, we will focus on two of the most common types of NoSQL databases: document-oriented databases and graph databases and explain the use cases suitable for each of them.

16:00 - 16:30 Coffee break

16:30 - 18:00 Multidisciplinary research and data analytics: Smart Cities (Maria Cristina Marinescu, Computer Applications in Science&Engineering, BSC)

A huge quantity of data is produced in cities from many types of sources: IoT, social network, other text sources, images, etc. Data integration is the first and more difficult step to ensure data quality and be able to then analyze these data and get insight that may help improve quality of life, sustainability, and resilience of the urban fabrics. This session focuses on the variety aspect of big data, and modeling as a way to capture common sense and enable semantic reasoning.

Day 3 (Feb 5)

9:30 – 13:00 Data Analytics with Apache Spark (Josep Lluís Berral, Computer Sciences - Data Centric Computing, BSC)

11:00 - 11:30 Coffee break

Apache Spark has become a consolidated technology for large-scale processing in a fast and

general way, with “programmer-friendly” interfaces and official bindings for many of the most used languages (Java, Scala, Python and R), extensive documentation and development tools. This course introduces Apache Spark, as well as some of its core libraries for data manipulation, machine learning, data streams and graph analytics.

13:00 – 14:00 Lunch Break

14:00 – 15:30 Data Analytics with Apache Spark. Part 2

16:00 – 16:15 Coffee break

15:30 – 17:00 European project on Big Data

Day 4 (Feb 6)

9:30 – 13:00 Practical Introduction to Python Deep Learning (Jordi Torres, Emerging Technologies for Artificial Intelligence Group Manager - Computer Sciences, BSC)

Artificial Intelligence is changing our lives, and solutions based on Deep Learning are leading this transformation. Deep Learning is now of major interest to companies and research centers, since it can be applied to many areas of activity. But getting started in this technology is not an easy task. The purpose of this short course is to gradually start the student off to the basics of Python Deep Learning, in a practical way through a guided, hands-on learning without becoming too technical, ensuring that the student learn enough of the basics to get literate in Deep Learning. Using the Keras API of TensorFlow library allows the development of Deep Learning models and abstracts much of the mathematical complexity involved in its implementation. The course content will be as follows:

PART 1: INTRODUCTION

1. What is Deep Learning?
2. Work environment
3. Python and its libraries

PART 2: FUNDAMENTALS OF DEEP LEARNING

4. Densely connected neural networks.
5. Neural networks in Keras
6. How a neural network is trained
7. Parameters and hyperparameters in neural networks
8. Convolutional neural networks.

PART 3: DEEP LEARNING TECHNIQUES

9. Stages of a Deep Learning project
10. Data to train neural networks
11. Data Augmentation and Transfer Learning
12. Advanced neural network architectures

PART 4: GENERATIVE DEEP LEARNING

Recurrent neural networks
Generative Adversarial Networks

Important prerequisites to enroll in this course: It is assumed that the student has a basic knowledge of Python prior to starting the course.

11:00 - 11:30 Coffee break

13:00 – 14:00 Lunch Break

14:00 - 16:00 From Data Mining to Data Science (Tomàs Aluja, UPC – Barcelona Tech)

Data contains information. We will try to contextualize the flow of apparently “new” concepts such as data mining, business intelligence, big data, data science and how they relate to the old school of exploratory statistics. We will also introduce an overview of the main steps of a data mining problem, and we will illustrate them through sound examples of application.

16:00 - 16:30 Coffee break

16:30 – 18:00 Data analytics in societal challenges modeling: smart mobility and other related fields (Dra. Mari Paz Linares i Jamie Arjona (UPC, inLab FIB)

Internet of Things, Big Data, Smart cities or Industry 4.0 are concepts that have raised in the last years with promises of solving daily human issues. In this session we will present how a combination of Internet of Things and Big Data can attack certain challenges and alleviate them.

Day 5 (Feb 7)

9:30 – 13:00 Data Visualization Theory (Luz Calvo, User Experience And Interaction Designer, BSC and Juan Felipe Gomez Celis, FrontEnd Developer, BSC)

Theory

Basic concepts

Human perception

Design

Colour

Audience / Validation / Bad practices

Visualisation design process

11:00 - 11:30 Coffee break

Tools for data visualization

– Tableau

– Data Wrapper

– RawGraphs

– Flourish

– Carto

Data visualisation with d3.js

END of COURSE