

Big Data for Engineers

SEMESTER: Spring
CREDITS: 3 ECTS
LANGUAGE: Spanish/English
DEGREES: MII

Course overview

This course takes you through the different technologies of Big Data, from the initial approach of data storage and processing, BI universe tools, to the dimensioning of a functional cluster.

By the end of the course, students will:

- Understand the basic principles behind distributed systems.
- Understand the pieces that make up the pillars of a Big Data architecture.
- Have a practical experience sizing a cluster based on use cases.
- Have well-formed criteria to choose how to deploy a Big Data cluster with the current technological tools and options.

Prerequisites

Students willing to take this course should be familiar with basic Linux commands and knowledge of computer architecture.

Course contents

Theory

1. Distributed systems
 - 1.1. Introduction to Datacenter
 - 1.2. Distributed processing
 - 1.3. Virtualization.
2. Hadoop
 - 2.1. Introduction
 - 2.2. HDFS
 - 2.3. YARN and Map Reduce
 - 2.4. Monitoring and administration
3. Commercial Big Data platforms
 - 3.1. On-premise
 - 3.2. Cloud computing
4. Design of Big Data architectures
 - 4.1. Methodology of design and sizing of a cluster.

Practice

The contents have been designed from a basic starting level, with the purpose of introducing participants to the technologies and tools most used in Big Data. Making an immersion to the most important concepts and demonstrations and practical exercises in each session.

Textbook

- Notes prepared by the lecturer (available in Moodle)
- Hadoop: The Definitive Guide, 4th Edition, ISBN-13: 978-1491901632
- Learning Spark: Lightning-Fast Big Data Analysis 1st Edition ISBN-13: 978-1491901632

Grading

- **Final exam** will account for 70%.
- **Lab** will account for the remaining 30%.
- In order to pass the course, the mark of the final exam must be greater or equal to 5 out of 10 points.