

DOI-MBD-523 Hybrid Architectures and Systems Administration

SEMESTER: Spring
CREDITS: 1,5 ECTS
LANGUAGE: Spanish/English
DEGREES: MBD

Course overview

This course will cover hybrid advanced analytics architectures, tools for distributed computing cluster management and advanced configurations to address real use cases.

By the end of the course, students will:

- They will be capable of responding to problems involving the storage and process of large amounts of data.
- Have well-formed criteria to choose the correct architecture to face uses cases of batch and real-time process and mixed scenarios.

Prerequisites

Students willing to take this course should be familiar with basic Hadoop commands and knowledge of the main Big Data tools.

Course contents

Theory

1. Hybrids Architectures.
 - 1.1. Lambda, Kappa, ...
 - 1.2. Elasticsearch y Solr
2. Advanced Administration
 - 2.1. Big Data Systems Administration: monitoring
 - 2.2. Oozie, Zookeeper, Cloudera Manager, Navigator, HUE
 - 2.3. Security
 - 2.4. Advanced Settings. Backup in Hadoop, Rackawareness and HA
 - 2.5. Advanced Use Cases

Practice

The contents have been designed from a medium level, with the objective of consolidating the concepts learned and continue to deepen the learning of advanced tools for the administration and deployment of Big Data clusters with demonstrations and practical exercises in each session.

Textbook

- Notes prepared by the lecturer
- Hadoop: The Definitive Guide, 4th Edition, ISBN-13: 978-1491901632
- Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine, ISBN: 1449358527, 9781449358525
- Questioning the Lambda Architecture. Jay Kreps. O'reilly

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.