

Subject

Big Data and Digital Ecosystem II

Year: 3

Credits: 8 ECTS

Language: Spanish

Competencies

Core competences:

CB1. Students have demonstrated knowledge and understanding in an area of study that builds on the foundation of general secondary education and is usually at a level that, while relying on advanced textbooks, also includes certain elements involving cutting-edge knowledge in their field of study

CB2. Students are able to apply their knowledge to their work or vocation in a professional manner and possess the skills that are typically demonstrated through the development and defence of arguments and problem solving within their field of study.

CB5. Students have developed those learning skills that are necessary to undertake further studies with a high degree of autonomy.

General competences:

CG4. Flexible and broad-minded Outlook

Specific competences:

CE1. Understand, know and apply the theories, tools and processes for capturing data sources from different types of data sources and for data warehousing

Learning outcomes

RA4. Ability to think and devise solutions and responses beyond the common and that stipulated by convention. Ability to respond to unexpected current and possible future circumstances by thinking outside the box, allowing them to understand the situation in which they find themselves.

RA8. Understand, know and apply the theories, tools and processes for the storage of data of a different type.

Syllabus

AWS:

- Introduction to cloud computing.
- Practice: Create EC2 machine.
- Basics:
 - EC2 compute layer:
 - Basics.
 - Template creation and deployment.
 - Guided exercises.
 - Practical exercises.
 - Data Centers: regions + availability zones
 - VPC network layer:
 - Basic concepts: subnets, gateways, network security (security groups).
 - Practical.
 - Storage layer:
 - Basics: S3 + CDN.
 - Exercises + practice.
 - Databases: RDS + NRDS
 - Practical exercises.

Hadoop Ecosystem:

- Distributed storage.
- Main components:
 - Data storage and management: HDFS (Distributed File System).
 - HDFS architecture: NameNode, DataNode, Secondary NameNode, MetaData.
 - Data processing: MapReduce (Simple and versatile programming paradigm for parallel programming).
 - Mapping phase and Reducing phase.
 - MapReduce architecture: JobTracker and TaskTracker.
 - Task scheduling and resource management: YARN resource manager.
- Replication.
- Sizing guide.
- Execution modes:
 - Standalone.
 - Pseudo-distributed.

- Cluster.
- Environment configuration and demonstration.

Apache Spark:

- Spark Core:
 - RRDs (Resilient Distributed Datasets):
 - Types of operations: transformations and actions.
 - Persistence.
 - Shared variables.
 - Programming model.
- Spark Application.
- SparkContext.
- SparkSession.
- Spark Clusters:
 - Local.
 - Standalone.
 - YARN.
 - MESOS.
- Spark SQL:
 - SQL Context.
 - DataFrames: schemas and operations.
 - SQL queries.

Introduction to Node-RED:

- Features and environment configuration.
- Main nodes.
- Programming model:
 - Multiple inputs and outputs.
 - Context module.
 - Subflows.
 - File reading and parsing.
- Node-RED Dashboard: active and passive nodes.

Introduction to Docker:

- Structure and functions:
 - Docker images.
 - Docker containers.
 - Docker Engine.
- Managing Docker containers and Docker images.
- Docker image creation:
 - Dockerfile.

Training activities

The training activities planned for this module are the following:

- Challenge-based learning (2 ECTS)
- Workshops (1 ECTS)
- Online resources (0.5 ECTS)
- Reflection (0.5 ECTS)
- Individual work (1 ECTS)
- Learning communities (1 ECTS)

Assessment system

Assessment will be by means of the continuous assessment system, providing constant feedback to both teachers and students on the learning process throughout the academic period:

- Learning activities involving the presentation of knowledge and individual study may be assessed by means of oral and/or written tests, which will account for a maximum of 60% of the final mark.
- The training activities aimed at acquiring the practical skills of the subjects will be assessed through the completion of various activities (assignments, case studies, challenges, etc.) accounting for at least 40% of the final mark.

Details of the assessment and marking will be made explicit in the annual academic planning of the subjects, in accordance with the teachers responsible and the determining factors of each course.

Bibliography

- Wittig, M., & Wittig, A. (2018). *Amazon web services in action*. Simon and Schuster.
- Morris, K. (2016). *Infrastructure as code: managing servers in the cloud*. "O'Reilly Media, Inc."