GUIA DOCENTE

MÁSTER UNIVERSITARIO EN DATA ANALYTICS FOR BUSINESS

Edición 2

Curso 2024-2025

1. ASIGNATURA

- **Nombre:** ***Fundamentos de Big Data***
- **Tipo de asignatura:** Obligatoria
- **Trimestre:** PRIMERO
- **Créditos:** 3 ECTS
- **Idioma de docencia:** castellano
- **Coordinador de la asignatura:** Luca Telloli
- **Datos de contacto**: luca.telloli+mudab@upf.edu
- **Profesor/es de la asignatura**: Luca Telloli
- **Datos de contacto:** luca.telloli+mudab@upf.edu

2. PRESENTACIÓN DE LA ASIGNATURA

El alumnado podrá entender los conceptos relacionados con big data mediante la aproximación a diferentes herramientas de almacenamiento y procesamiento de grandes cantidades de datos, tales como Apache Spark, disponibles en plataformas de procesamiento en la nube como Amazon Web Services.

Objetivos de la asignatura

- Adquirir conceptos y herramientas básicas en el ámbito del procesamiento de datos a gran escala
- Adquirir experiencia con herramientas de Big Data en la nube

Nuestro compromiso con el impacto social y el bienestar planetario se traduce en contenidos formativos alineados con los Objetivos de Desarrollo Sostenible (ODS) previstos en la agenda 2030:

En la asignatura que nos ocupa, los ODS implicados son:

- ODS.4. Educación de Calidad
- ODS.8. Trabajo Decente y Crecimiento Económico

Contenidos

- Introducción a Big Data. Datos estructurados y intercambio de datos. El formato JSON.
- Almacenamiento de Big Data.
- Procesamiento de Big Data. Map-Reduce.
- Big Data en la nube. AWS S3.
- Procesamiento en la nube. AWS EMR.
- Bases de datos NoSQL. Dynamo DB.

La asignatura dentro del plan de estudios

*Esta asignatura **obligatoria** se enmarca dentro de la materia 1. **Análisis de Datos. Data Analytics** del plan de estudios. Se realiza durante el **segundo trimestre.***

Competencias/Resultados de aprendizaje

RA1. Mat 1.1 Seleccionará la infraestructura necesaria para hacer frente a un proyecto que involucre datos masivos.

RA3 Mat 1.2 Realizará un informe completo que incluya visualizaciones diversas sobre un conjunto de datos y que ayuden a la toma de decisiones.

RA4. Mat 1.3 Extraerá información de modo visual a partir de datos masivos.

RA5. Mat. 1.4 Identificará las diferencias fundamentales entre bases de datos relacionales y no relacionales.

RA6. Mat 1.5 Analizará un conjunto de datos mediante análisis univariante y bivariante.

RA6. Mat 1.6 Distinguirá correlaciones entre las diferentes dimensiones de un conjunto de datos.

RA7. Mat 1.7 Realizará un plan de gobernanza de datos que minimice riesgos y costes.

RA19.Mat 1.1 Utilizará python para generar visualizaciones adecuadas al tipo de datos que se estén trabajando en diversos sectores empresariales

RA14. Mat 1.1 Planteará unas hipótesis sobre un conjunto de datos realizando un test de hipótesis que te permita saber si debes aceptar o rechazar la hipótesis nula.

RA14. Mat 1.2 Propondrá un algoritmo para predecir la aceptación, por parte de clientes banco visionarios, de diferentes campañas.

RA14. Mat 1.3 Construirá un algoritmo de clasificación de pacientes sanos vs enfermos.

RA14. Mat 1.4 Diseñará un algoritmo para la segmentación de clientes.

RA14. Mat 1.5 Lista todas las variables que deberías tener en cuenta para optimizar los procesos de almacenaje de una empresa.

RA14. Mat 1.6 Formulará nuevas aplicaciones que podría desarrollar en el ámbito empresarial o sectorial utilizando las herramientas más punteras de analítica de datos.

PLAN DE APRENDIZAJE DE LA ASIGNATURA

Metodología docente

Clases "hands-on": presentación de un problema e implementación a través de mejoras sucesivas.

Horas de dedicación (horas lectivas + trabajo del alumno): 75

Evaluación (sistema de evaluación, sistema de cualificación…)

- Participación en clase (10%)

- Ejercicios con entrega a través de Moodle (40%)
- Examen final a respuesta múltiple (50%)

upf. BARCELONA SCHOOL OF MANAGEMENT

Información sobre las sesiones

Primera sesión	Data and big data Lifecycle of data & big data Use cases for big data Data types: structured, semi and unstructured data. Schemas JSON & JSON parsing **The problem**: parsing semi-structured data with Python
Segunda sesión	- Storage of big data: replication, sharding - CAP theorem - **The problem**: parse a large collection of tweets locally
Tercera sesión	Processing of big data Batch versus real-time ETL: Extract – Transform – Load Map-reduce & the classic word count example **The problem**: sentiment analysis with map-reduce
Cuarta sesión	The cloud: Intro to AWS. Multi-modal access High-availability storage in the cloud: AWS S3 The problem: storing data in S3
Quinta sesión	The cloud: large-scale processing with AWS EMR The problem: processing data in EMR
Sexta sesión	NoSQL. DynamoDB.

3. PROFESORADO

Luca Telloli es ingeniero de datos en Adevinta, una de las mayores empresas de clasificados online. Titulado en Informática por la Universidad de Bologna y con un Máster en Seguridad Informática realizado entre Bologna y Estados Unidos, divide su tiempo entre su actividad primaria de ingeniero y la enseñanza.

Es docente en el departamento de Tecnologías de la Información (DTIC) de la Universidad Pompeu Fabra, donde ha impartido cursos de Sistemas Operativos, Aplicaciones Telemáticas, y donde actualmente imparte el curso de Sistemas Distribuidos a Gran Escala.

4. BIBLIOGRAFIA (obligatoria/ recomendada)

Recomendaciones:

- Erl, Khattak, and Buhler: *Big Data Fundamentals: Concepts, Drivers & Techniques*, Pearson, 2016
- Warren, Marz: Big Data. Manning Publications, 2015
- Kleppmann: Designing Data-Intensive Applications, O'Reilly Media, Inc., 2017
- Chambers, Zaharia: Spark: The Definitive Guide, O'Reilly Media, Inc., 2018