

**Proyecto/Guía docente de la asignatura**

Asignatura	TECNOLOGÍA PARA EL BIG DATA		
Materia	BIG DATA		
Módulo	TECNOLOGÍAS INFORMÁTICAS DEL BIG DATA		
Titulación	MÁSTER EN INGENIERÍA INFORMÁTICA		
Plan	510	Código	53203
Periodo de impartición	1 ^{er} CUATRIMESTRE	Tipo/Carácter	OPTATIVA
Nivel/Ciclo	MÁSTER	Curso	2º
Créditos ECTS	6 ECTS		
Lengua en que se imparte	CASTELLANO		
Profesor/es responsable/s	CÉSAR GONZÁLEZ FERRERAS IVÁN SANTOS TEJIDO		
Datos de contacto (E-mail, teléfono...)	TELÉFONO: 983 185622 - E-MAIL: cesargf@infor.uva.es TELÉFONO: 983 423000 ext. 5512 - E-MAIL: ivasan@tel.uva.es		
Departamento	DEPARTAMENTO DE INFORMÁTICA DEPARTAMENTO DE ELECTRÓNICA		

1. Situación / Sentido de la Asignatura

1.1 Contextualización

Esta asignatura pretende introducir las diversas tecnologías que se utilizan en las aplicaciones de análisis de grandes repositorios de datos. Se profundizará en las herramientas de análisis distribuido de datos más utilizadas. Por otro lado, se estudiará el proceso de puesta en marcha de la infraestructura necesaria para este tipo de aplicaciones. Se presentarán las diferentes plataformas de computación en la nube existentes, y se analizarán las diversas posibilidades que ofrecen para las aplicaciones big data.

1.2 Relación con otras materias

Esta asignatura se relaciona con conceptos presentados en las siguientes asignaturas:

- Técnicas Escalables de Análisis de Datos
- Almacenamiento Escalable
- Big Data: inteligencia de negocios

1.3 Prerrequisitos

2. Competencias

2.1 Generales

2.2 Específicas

Código	Descripción
CET1	Capacidad para modelar, diseñar, definir la arquitectura, implantar, gestionar, operar, administrar y mantener aplicaciones, redes, sistemas, servicios y contenidos informáticos.
CET5	Capacidad para analizar las necesidades de información que se plantean en un entorno y llevar a cabo en todas sus etapas el proceso de construcción de un sistema de información.
CET6	Capacidad para diseñar y evaluar sistemas operativos y servidores, y aplicaciones y sistemas basados en computación distribuida.
CET9	Capacidad para aplicar métodos matemáticos, estadísticos y de inteligencia artificial para modelar, diseñar y desarrollar aplicaciones, servicios, sistemas inteligentes y sistemas basados en el conocimiento.
CBD1	Capacidad de aplicar técnicas de análisis escalable de datos en diferentes ámbitos de actuación utilizando las tecnologías adecuadas.
CBD4	Capacidad de implementar sistemas de descubrimiento de conocimiento en grandes bases de datos distribuidas.
CBD5	Capacidad de analizar, diseñar y construir o configurar sistemas de almacenamiento escalable y procesamiento escalable

3. Objetivos

Código	Descripción
CBD1.1	Conocer los principales lenguajes de programación utilizados en Big Data y saber programar en detalle con al menos uno de ellos.
CBD4.1	Conocer las herramientas de análisis distribuido de datos que existen junto con sus principales características. Saber manejar y resolver problemas sencillos con alguna de las

	diferentes herramientas de análisis distribuido de datos.
CBD5.1	Conocer los diferentes elementos software que se pueden utilizar para configurar una infraestructura dedicada al Big Data. Saber crear y administrar una infraestructura sencilla de Big Data.
CBD5.2	Conocer las diferentes plataformas de computación en la nube existentes junto con sus características. Desarrollar y configurar proyectos de Big Data utilizando dichas plataformas.

4. Contenidos y/o bloques temáticos

Bloque 1: Lenguajes y Herramientas para el Big Data

Carga de trabajo en créditos ECTS:

a. Contextualización y justificación

En este bloque se trabajan las competencias relacionadas con los lenguajes y las herramientas orientados al big data.

b. Objetivos de aprendizaje

Código	Descripción
CBD1.1	Conocer los principales lenguajes de programación utilizados en Big Data y saber programar en detalle con al menos uno de ellos.
CBD4.1	Conocer las herramientas de análisis distribuido de datos que existen junto con sus principales características. Saber manejar y resolver problemas sencillos con alguna de las diferentes herramientas de análisis distribuido de datos.

c. Contenidos

TEMA 1: Lenguajes para el Big Data

TEMA 2: Introducción a la minería de datos

TEMA 3: Modelo de Programación MapReduce

TEMA 4: Herramientas para análisis distribuido

d. Métodos docentes

Ver apartado 5: Métodos docentes y principios metodológicos

e. Plan de trabajo

Se desarrollarán dos entregas prácticas empleando las tecnologías desarrolladas en la asignatura.

f. Evaluación

Ver apartado 7: sistema y características de la evaluación.

g. Bibliografía básica

"Hadoop: The Definitive Guide". Tom White. O'Reilly Media / Yahoo Press. 2012

"Spark in action". Petar Zecevic, Marko Bonaci. Manning, 2017.

"Learning Spark: Lightning-Fast Big Data Analysis" Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia. O'Reilly 2015.

"Mining of Massive Datasets". Jure Leskovec, Anand Rajaraman, Jeff Ullman (<http://www.mmids.org/>)

"Python for Data Analysis". Wes McKinney. O'Reilly Media, 2012.

"Python Data Analysis". Ivan Idris. Packt Publishing, 2014

h. Bibliografía complementaria

i. Recursos necesarios

- El alumno deberá tener acceso a un ordenador personal.
- Aula virtual de la asignatura.
- Software:
 - o Hadoop
 - o Spark

j. Temporalización

CARGA ECTS	PERIODO PREVISTO DE DESARROLLO
4 ECTS	Semanas 1 a 5

Bloque 2: Infraestructuras y Plataformas para el Big Data

Carga de trabajo en créditos ECTS:

a. Contextualización y justificación

En este bloque se trabajan las competencias relacionadas con las infraestructuras y las plataformas para las aplicaciones de big data.

b. Objetivos de aprendizaje

Código	Descripción
CBD5.1	Conocer los diferentes elementos software que se pueden utilizar para configurar una infraestructura dedicada al Big Data. Saber crear y administrar una infraestructura sencilla de Big Data.
CBD5.2	Conocer las diferentes plataformas de computación en la nube existentes junto con sus características. Desarrollar y configurar proyectos de Big Data utilizando dichas plataformas.

c. Contenidos**TEMA 5: Infraestructura para el Big Data****TEMA 6: Plataformas en la nube****d. Métodos docentes**

Ver apartado 5: Métodos docentes y principios metodológicos

e. Plan de trabajo

Ver Cronograma de actividades previstas

f. Evaluación

Ver apartado 7: sistema y características de la evaluación.

g. Bibliografía básica

"Big Data. Related Technologies, Challenges and Future Prospects". Min Chen, Shiwen Mao, Yin Zhang, Víctor C.M. Leung. Springer 2014

"High-Performance Big-Data Analytics. Computing Systems and Approaches". Pethuru Raj, Anupama Raman, Dhivya Nagaraj, Siddhartha Duggirala. Springer 2015

"Cloud Networking for Big Data". Deze Zeng, Lin Gu, Song Guo. Springer 2015

h. Bibliografía complementaria

"Big Data Made Easy. A Working Guide to the Complete Hadoop Toolset". Michael Frampton. Apress 2015

"Big Data Analytics Infrastructure For Dummies" Barry Schoenborn. John Wiley & Sons 2014

i. Recursos necesarios

El alumno deberá tener acceso a un ordenador personal.

El profesor proporcionará los recursos necesarios para realizar este bloque a través del Campus Virtual.

j. Temporalización

CARGA ECTS	PERIODO PREVISTO DE DESARROLLO
2 ECTS	Semanas 6 a 8

5. Métodos docentes y principios metodológicos

Actividad	Metodología
Clase de teoría	<ul style="list-style-type: none">• Clase magistral participativa• Estudio de casos en aula• Resolución de problemas
Clase práctica	<ul style="list-style-type: none">• Realización de un trabajo práctico guiado por el profesor.• Clase magistral participativa

	<ul style="list-style-type: none">Resolución de casos prácticos con apoyo informáticoRealización de un proyecto en grupo (2/3 alumnos) guiado por el profesor, siguiendo un enfoque colaborativo.
Seminarios	<ul style="list-style-type: none">Talleres de aprendizajeSesiones de debate entre alumnos y profesor sobre su aprendizaje, las técnicas estudiadas y su aplicación práctica a casos reales.

6. Tabla de dedicación del estudiante a la asignatura

ACTIVIDADES PRESENCIALES	HORAS	ACTIVIDADES NO PRESENCIALES	HORAS
Clases teórico-prácticas (T/M)	26	Estudio y trabajo autónomo individual	45
Clases prácticas de aula (A)		Estudio y trabajo autónomo grupal	45
Laboratorios (L)	30		
Prácticas externas, clínicas o de campo			
Seminarios (S)	4		
Tutorías grupales (TG)			
Evaluación (fuera del periodo oficial de exámenes)			
Total presencial	60	Total no presencial	90

7. Sistema y características de la evaluación

INSTRUMENTO/PROCEDIMIENTO	PESO EN LA NOTA FINAL	OBSERVACIONES
Bloque 1: entrega y defensa de la práctica 1	1/3	Semana 3
Bloque 1: entrega y defensa de la práctica 2	1/3	Semana 5
Bloque 2: entrega y defensa de la práctica 3	1/3	Semana 8

CRITERIOS DE CALIFICACIÓN

Convocatoria ordinaria:

- Para poder aprobar la asignatura es necesario sacar más de un 4 sobre 10 en cada una de las tres prácticas de la asignatura, y más de un 5 en el promedio de la nota de las prácticas.

Convocatoria extraordinaria:

- Para la realización de las prácticas será necesario contactar con el profesor y fijar una planificación para el desarrollo de las mismas. En todo caso, será necesario entregar y defender tres trabajos prácticos.
- Para poder aprobar la asignatura es necesario sacar más de un 4 sobre 10 en cada una de las tres prácticas de la asignatura, y más de un 5 en el promedio de la nota de las prácticas.

8. Consideraciones finales