



## FICHA TÉCNICA DE LA ASIGNATURA

<b>Datos de la asignatura</b>	
<b>Nombre completo</b>	Datos Masivos / Big Data
<b>Código</b>	E000008337
<b>Impartido en</b>	Grado en Análisis de Negocios/Business Analytics y Grado en Derecho [Tercer Curso]
<b>Nivel</b>	Reglada Grado Europeo
<b>Cuatrimestre</b>	Semestral
<b>Créditos</b>	3,0 ECTS
<b>Carácter</b>	Obligatoria (Grado)
<b>Departamento / Área</b>	Departamento de Telemática y Computación
<b>Responsable</b>	Carlos Miguel Vallez Fernández
<b>Horario de tutorías</b>	Petición via email

<b>Datos del profesorado</b>	
<b>Profesor</b>	
<b>Nombre</b>	Carlos Miguel Vallez Fernández
<b>Departamento / Área</b>	Departamento de Telemática y Computación
<b>Correo electrónico</b>	cmvallez@icai.comillas.edu

## DATOS ESPECÍFICOS DE LA ASIGNATURA

<b>Contextualización de la asignatura</b>
<b>Aportación al perfil profesional de la titulación</b>
<p>La tecnología Big Data es esencial en la disciplina de Analítica avanzada y procesamiento masivo de datos y una herramienta de uso cada vez más extendido en las empresas como habilitador de los análisis avanzados. De esta manera el objetivo de la asignatura es que el alumno aprenda los conceptos relacionados de Big data, desde sus origen e historia, características y diferencias respecto a sistemas tradicionales, su arquitectura paralela y las diferencias con otros sistemas paralelos y ecosistema del ciclo del dato básico, desde su adquisición del dato, almacenamiento al procesamiento. Se relacionará las capacidades que ofrece esta tecnología con los problemas que para los que esta especialmente indicada, las necesidades que cubre y los casos de uso que consiguen aportar valor diferencial a las empresas y la sociedad.</p> <p>Al finalizar el curso, los alumnos entenderán los bloques habituales de una instalación tipo, el funcionamiento de big data, su utilidad y sabrán en qué casos debe ser aplicado, por qué y para qué, de tal manera que puedan proponer y defender casos concretos de uso.</p>



## Prerequisitos

Ninguno

## Competencias - Objetivos

### Competencias

#### Competencias Básicas

CB 02. Saber aplicar e integrar sus conocimientos, la comprensión de éstos, su fundamentación científica y sus capacidades de resolución de problemas en entornos nuevos y definidos de forma imprecisa, incluyendo contextos de carácter multidisciplinar tanto investigadores como profesionales altamente especializados

CB 03. Saber evaluar y seleccionar la teoría científica adecuada y la metodología precisa de sus campos de estudio para formular juicios a partir de información incompleta o limitada incluyendo, cuando sea preciso y pertinente, una reflexión sobre la responsabilidad social o ética ligada a la solución que se proponga en cada caso.

CB 07. Ser capaces de asumir la responsabilidad de su propio desarrollo profesional y de su especialización en uno o más campos de estudio.

#### Competencias Generales

CG. 1 Capacidad de organización y planificación en la identificación de problemas en el contexto de datos masivos

CG 02. Capacidad de análisis de datos masivos procedentes de diversas fuentes: texto, audio, numérica e imagen.

CG 03. Resolución de problemas y toma de decisiones en un entorno de datos masivos tanto cuantitativos como cualitativos.

CG 07. Capacidad de liderazgo y trabajo en equipo en la sociedad de la información.

CG 8. Capacidad crítica y autocrítica en la sociedad de la información.

CG 9. Compromiso ético en la sociedad de la información.

CG 11. Capacidad para aprender y trabajar autónomamente en la sociedad de la información.

#### Competencias Específicas

CE 25. Comprender los conceptos más importantes asociados al Big Data y sus elementos más característicos.

#### Competencias Específicas de la Optatividad

CEO 24. Conocer y usar técnicas de análisis de información para extraer el conocimiento para la toma



de decisiones de negocio.

### Resultados de Aprendizaje

#### **CE 25. Comprender los conceptos más importantes asociados al Big Data y sus elementos más característicos**

RA1. Conocer el concepto de paralelismo y la importancia de la distribución de datos voluminosos.

RA2. Identificar las necesidades específicas asociadas al manejo, adquisición y almacenamiento de volúmenes masivos de datos.

RA3. Comprender y conocer las tecnologías de comunicaciones (redes) y computación precisas para la distribución y recuperación de información, así como de su análisis.

#### **CE 24. Conocer y usar técnicas de análisis de información para extraer el conocimiento para la toma de decisiones de negocio.**

RA1 Conoce y comprende las herramientas de apoyo a la decisión en el ámbito de la empresa.

RA2 Es capaz de implementar una solución computerizada de apoyo a la decisión.

RA3 Comprende y aplica herramientas de adquisición, representación e inferencia, y aquellas técnicas de aprendizaje automático y basadas en sistemas inteligentes.

## BLOQUES TEMÁTICOS Y CONTENIDOS

### Contenidos – Bloques Temáticos

#### **BLOQUE 1: Contenido de la asignatura**

##### **1: Introducción a Big Data**

1. ¿Qué es Big Data y qué no es Big Data?
2. El valor del dato
3. Volúmenes y fuentes de datos
4. Información estructurada y no estructurada
5. Tecnologías exponenciales y tecnologías disruptivas

##### **2: Sistemas Distribuidos**

1. Introducción
2. Definición y características



3. Problemas. Tolerancia a fallos
4. Balanceo de carga de servidores
5. Virtualización:
  - Conceptos y tipos de virtualización
  - Máquinas virtuales vs contenedores
  - Sistemas operativos Linux

### **3: Ecosistema Hadoop**

1. Necesidad del ecosistema
2. Esquema
3. Introducción a los diversos componentes
4. Sistema de ficheros HDFS
5. Adquisición de datos. Scoop, Flume y Kaffka
6. Búsqueda y procesamiento de datos. Hbase y Hive
7. Análisis de datos
8. Arquitectura básica, Administración y monitorización de un cluster

### **4: Procesamiento Masivo**

1. YARN
2. MapReduce
3. Spark

### **5: Cloud & Big Data**

1. Definición de la nube
2. Concepto "as a service"
3. Proveedores Cloud

### **6: Tecnología Big Data mediante casos de uso**

1. Impacto de Big Data en Analytics y la Inteligencia Artificial/Machine Learning
2. Palancas de uso de Big Data empresarial
3. Casos de uso de tecnología Big Data

## **BLOQUE 2: Prácticas de la asignatura**

1. Redistribución y unificación de servidores. Redundancia, tolerancia a fallos y distribución de recursos.
2. Instalación software de virtualización y puesta en marcha de una máquina virtual Cloudera.
3. Ejercicios de comandos Linux
4. Práctica guía de ejemplos Map-reduce
5. Elaboración de un caso de uso en el cual se aplique Map Reduce
6. HDFS: Sistema de Almacenamiento en Hadoop
7. Ejercicios y comparativa de ejecuciones con Spark
8. Cuadro comparativo de características de la oferta de los principales proveedores de servicios cloud.
9. Contratación y puesta en marcha de un servicio cloud
10. Uso de Jupyter notebook y herramienta comercial para el análisis de datos, visualización obtención de predicciones.



## METODOLOGÍA DOCENTE

### Aspectos metodológicos generales de la asignatura

#### Metodología Presencial: Actividades

1. **Clase magistral y presentaciones generales.** Exposición de los principales conceptos y procedimientos mediante la explicación por parte del profesor. Incluirá presentaciones dinámicas, pequeños ejemplos prácticos y la participación reglada o espontánea de los estudiantes. Además, el profesor propondrá la lectura de temas preparados por el profesor, artículos científicos y divulgativos que permitirán introducir o profundizar en un tema, con el objetivo de aportar ideas y generar debate **(24 horas).CE 25,CG 03, CG 02 CG 01, CB 03, CB 02**
2. **Resolución en clase de problemas prácticos.** Resolución de unos primeros problemas para situar al alumno en contexto. La resolución correrá a cargo del profesor y los alumnos de forma cooperativa **(2 horas). CE 25,CG 03, CG 02 CG 01, CB 03, CB 02**
3. **Prácticas.** Cada alumno realizará de forma aislada o en grupo una serie de prácticas regladas. Las prácticas finalizarán con la redacción de un informe o la redacción de las distintas experiencias **(4 horas). CE 25,CG 03, CG 02 CG 01, CB 03, CB 02**

#### Metodología No presencial: Actividades

1. **Estudio individual del material.** Actividad realizada individualmente por el estudiante para preparar el tema que se discutirá en clase. **(60 horas). CE 25,CG 03, CG 02 CG 01, CB 03, CB 02**
2. **Resolución de problemas prácticos** a resolver fuera del horario de clase por parte del alumno. El alumno debe utilizar e interiorizar los conocimientos aportados en la materia. La corrección a la clase se realizará por parte de alguno de los alumnos o el profesor según los casos. La corrección individualizada de cada ejercicio la realizará el propio alumno u otro compañero según los casos (método de intercambio) **(10 horas).CE 25,CG 03, CG 02 CG 01, CB 03, CB 02**
3. **Prácticas de la asignatura.** Realización de dos prácticas que plasme los conocimientos adquiridos en la asignatura. **(10 horas).CE 25,CG 03, CG 02 CG 01, CB 03, CB 02**

## RESUMEN HORAS DE TRABAJO DEL ALUMNO

### RESUMEN HORAS DE TRABAJO DEL ALUMNO

#### HORAS PRESENCIALES

Lección magistral 24

Debates y resolución de problemas 2

Prácticas 4

Evaluación 2

#### HORAS NO PRESENCIALES



**Trabajo autónomo sobre contenidos teóricos 10**

**Trabajo autónomo sobre contenidos prácticos 10**

**Práctica final de la asignatura 10**

**Estudio 30**

**CRÉDITOS ECTS:**

**3 (90 horas)**

## **EVALUACIÓN Y CRITERIOS DE CALIFICACIÓN**

### **Actividades de evaluación**

#### **Criterios de evaluación**

**PESO**

#### **Realización de exámenes:**

- Examen final.
- Comprensión de conceptos mediante pruebas abiertas, enfocado a casos prácticos.
- Aplicación de los conceptos adquiridos
- Presentación y comunicación escrita.

**60%**

#### **Evaluación del rendimiento teórico continuo:**

Realización de pruebas de seguimiento, debates y ejercicios.

- Comprensión de conceptos mediante pruebas objetivas y one minute papers.
- Debate y análisis de los documentos técnico/científicos.
- Análisis e interpretación de los resultados obtenidos en la resolución de problemas mediante debates.

**30%**

#### **Evaluación del rendimiento práctico continuo:**

- Casos a desarrollar en la clase de prácticas.
- Comprensión de conceptos.
- Aplicación de conceptos al desarrollo de casos sobre los nuevos conceptos planteados.
- Aplicación de todos los conocimientos adquiridos durante el curso.

**10%**



## Calificaciones

### Criterios de Calificación

La calificación en la **convocatoria ordinaria** de la asignatura se obtendrá como:

- Un 60% la calificación del examen final.
- Un 30% será la calificación de las pruebas prácticas
  - 10% corresponde a prácticas grupales
  - 10% prácticas individuales
  - 10% breves test- exámenes- entregas de prácticas guiadas realizadas en clase.
- Un 10% será la calificación de la asistencia y participación en clase.

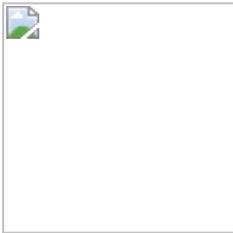
La calificación en la **convocatoria extraordinaria**:

- Un 60% la calificación del examen final.
- Un 30% será la calificación de las pruebas prácticas
  - 10% corresponde a prácticas grupales
  - 10% prácticas individuales
  - 10% breves test- exámenes- entregas de prácticas guiadas realizadas en clase.
- Un 10% será la calificación de la asistencia y participación en clase.

**Para aprobar la asignatura los alumnos tienen que tener al menos 5 puntos sobre 10 en el examen final, y en las práctica, tanto en la convocatoria ordinaria como en la extraordinaria.**

## PLAN DE TRABAJO Y CRONOGRAMA

Actividades	Fecha de realización	Fecha de entrega
<p><b>Actividades Presenciales y No presenciales</b></p> <p><b>Fecha de realización</b></p> <p><b>Fecha de entrega</b></p> <ul style="list-style-type: none"><li>• Lectura y estudio de los contenidos teóricos Después de cada clase</li><li>• Resolución de los problemas propuestos o debates Semanalmente</li></ul>		



<ul style="list-style-type: none"><li>• Preparación de las prácticas</li></ul> <p>Semanas</p> <p>8, 9 y 10</p> <p>Semanas</p> <p>11</p> <ul style="list-style-type: none"><li>• Preparación del examen final</li></ul>		
--	--	--

## **BIBLIOGRAFÍA Y RECURSOS**

### **Bibliografía Básica**

- Notas preparadas por el profesor disponibles en Moodle.

### **Bibliografía Complementaria**

- Big Data: A Revolution That Will Transform How We Live, Work, and Think. [Viktor Mayer-Schonberger](#) (Autor), [Kenneth Cukier](#) (Autor)
- George Orwell. 1984 . ISBN: 9788499890944
- Michael Lewis. Moneyball: The Art of Winning an Unfair Game Paperback – March 17, 2004. ISBN: 978-0393324815 or the film 'Moneyball (2011)'. Sony Pictures Director: Bennett Miller