

FICHA TÉCNICA DE LA ASIGNATURA

Datos de la asignatura	
Nombre completo	Datos Masivos / Big Data
Código	E000008337
Impartido en	Grado en Análisis de Negocios/Business Analytics y Grado en Derecho [Tercer Curso] Grado en Administración y Dirección de Empresas y Grado en Análisis de Negocios/Business Analytics [Tercer Curso]
Nivel	Reglada Grado Europeo
Cuatrimestre	Semestral
Créditos	3,0 ECTS
Carácter	Obligatoria (Grado)
Departamento / Área	Departamento de Telemática y Computación
Responsable	Carlos Miguel Vallez Fernández
Horario	Revisar Horario
Horario de tutorías	Concertar cita con el profesor

Datos del profesorado	
Profesor	
Nombre	Carlos Miguel Vallez Fernández
Departamento / Área	Departamento de Telemática y Computación
Despacho	Se comunicará al comienzo del curso
Correo electrónico	cmvallez@icai.comillas.edu

DATOS ESPECÍFICOS DE LA ASIGNATURA

Contextualización de la asignatura
<p>Aportación al perfil profesional de la titulación</p> <p>La tecnología Big Data es esencial en la disciplina de Analítica avanzada y procesado 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



Conocimientos básicos de programación.

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

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. Información estructurada y no estructurada.
4. Utilidad, necesidad y ejemplos de casos de uso Big Data.

2: Ciclo de vida de un proyecto Big Data

1. Data Pipeline de un proyecto Big Data. (Ingestión, almacenamiento, procesado/computación y Presentación/visualización).
2. Perfiles profesionales.
3. Ejemplo básico de ingestión y problemática asociada por recursos disponibles.
4. Introducción a Gobierno, privacidad y protección del dato. Perfil DPO , RGPD y derechos ARCO.

3: Conceptos básicos sobre sistemas

1. Sistemas operativos basados en Linux. Introducción y comandos elementales.
2. Sistemas Distribuidos. Introducción y conceptos relacionados (Paralelización, Tolerancia a fallos, Balanceo, Disponibilidad, Redundancia)
3. Concepto de virtualización. Introducción y plataformas. Máquinas virtuales vs contenedores.
4. Datacenters: redes de alta velocidad y computación de alto rendimiento.

4: Ecosistema Hadoop

1. Introducción a Hadoop. Componentes
2. Arquitectura Hardware y Software.

5: Almacenamiento Masivo

1. Sistema de ficheros HDFS.
2. Servicios y roles.

3. HUE. Recuperación y visualización de datos.

6: Introducción al Procesamiento Masivo

1. YARN.
2. MapReduce.
3. Spark.

7: Cloud & Big Data

1. Infraestructura on-premise vs Cloud.
2. Concepto "as a service".
3. Proveedores y plataformas Cloud. Servicios y productos ofrecidos.

8: Tecnología Big Data en la empresa

1. Impacto de Big Data en Analytics y la Inteligencia Artificial/Machine Learning.
2. Big Data como tecnología exponencial y disruptiva en el negocio.

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 9

Trabajo autónomo sobre contenidos prácticos 9

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:

- Examen final (entre 50.0 % y 60.0%)
 - **Examen final con un peso del 50%**
- Exámenes breves de contenido teórico o práctico, de desarrollo o tipo test. (entre 10.0 % y 30.0%)
 - **15% Exámenes Breves estilo cuestionario durante las clases**
- Evaluación de trabajos individuales (entre 10.0 % y 10.0%)
 - **10% Informes de prácticas individuales**
- Evaluación de trabajos en grupo (entre 10.0 % y 20.0%)
 - **15% Informes de prácticas grupales**
- Participación activa del alumno en el aula (entre 10.0 % y 20.0%)
 - **10% calificación de la asistencia y participación en clase**

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

- Examen final (entre 50.0 % y 60.0%)
 - **Examen final con un peso del 60%**
- Exámenes breves de contenido teórico o práctico, de desarrollo o tipo test. (entre 10.0 % y 30.0%)
 - **10% Exámenes Breves estilo cuestionario durante las clases**
- Evaluación de trabajos individuales (entre 10.0 % y 10.0%)
 - **10% Informes de prácticas individuales**
- Evaluación de trabajos en grupo (entre 10.0 % y 20.0%)
 - **10% Informes de prácticas grupales**
- Participación activa del alumno en el aula (entre 10.0 % y 20.0%)
 - **10% calificación de la asistencia y participación en clase**

Para aprobar la asignatura los alumnos tienen que tener al menos 5 puntos sobre 10 tanto en el examen final como en las prácticas. Esto aplica tanto en la convocatoria ordinaria como en la extraordinaria.

PLAN DE TRABAJO Y CRONOGRAMA

Actividades	Fecha de realización	Fecha de entrega



Actividades Presenciales y No presenciales

Fecha de
realización

Fecha de
entrega

- Lectura y estudio de los contenidos teóricos

Después de cada clase

- Resolución de los problemas propuestos o debates

Semanalmente

- Preparación de las prácticas

Semanas

8, 9 y 10

Semanas

11

- Preparación del examen final

BIBLIOGRAFÍA Y RECURSOS

Bibliografía Básica

- Notas preparadas por el profesor disponibles en Moodle.

Bibliografía Complementaria

- Mayer-Schönberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.
- P. Zikopoulos, C. Eaton, D. DeRoos, T. Deutsch, & G. Lapis, (2012) Understanding Big Data, USA: McGraw-Hill Books
- White, T. (2009). Hadoop. The Definitive Guide. Vol. 1, O'Reilly Media. Inc, San Jose, CA, USA
- L. Joyanes (2013), Big Data: Análisis de grandes volúmenes de datos en organizaciones, Editorial Alfaomega
- Chambers, B., & Zaharia, M. (2018). Spark: The definitive guide: Big data processing made simple. " O'Reilly Media
- Holmes, A. (2012). Hadoop in practice. Manning Publications Co..
- Shotts, W. (2019). The Linux command line: a complete introduction. No Starch Press.
- Karau, H., Konwinski, A., Wendell, P., & Zaharia, M. (2015). Learning spark: lightning-fast big data analysis. " O'Reilly Media, Inc."



COMILLAS

UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

GUÍA DOCENTE

2021 - 2022

- M. van Steen and A.S. Tanenbaum, Distributed Systems, 3rd ed., distributed-systems.net, 2017.
- Noah Iliinsky and Julie Steele Data Visualizations O'Reilly
- Nathan Yau (2011) Visualize This: The FlowingData Guide to Design, Visualization, and Statistics
- Lutz, M. (2009). Python Pocket Reference: Python in Your Pocket" O'Reilly Media, Inc."
- Lutz, M. (2013). Learning python: Powerful object-oriented programming. " O'Reilly Media, Inc."