

FICHA TÉCNICA DE LA ASIGNATURA

Datos de la asignatura		
Nombre completo	Datos Masivos / Big Data	
Código	DTC-BA-423	
Título	Grado en Análisis de Negocios / Business Analytics por la Universidad Pontificia Comillas	
Nivel	Reglada Grado Europeo	
Cuatrimestre	Semestral	
Créditos	3,0 ECTS	
Carácter	Obligatoria (Grado)	
Departamento / Área	Departamento de Telemática y Computación	
Responsable	Moisés Martínez	
Horario	Consultar horarios de la escuela	
Horario de tutorías	Concertar cita con el profesor	

Datos del profesorado		
Profesor		
Nombre	Pablo Martínez-Almeida Ramos	
Departamento / Área	Departamento de Telemática y Computación	
Correo electrónico pmartinezalmeida@icai.comillas.edu		

DATOS ESPECÍFICOS DE LA ASIGNATURA

Contextualización de la asignatura

Aportación al perfil profesional de la titulación

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.

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.

Prerrequisitos

Conocimientos básicos de programación.

Competencias - Objetivos



Competencia	ıs	
GENERALES		
CG01	Capacidad de organización y planificación en la identificación de problemas en el contexto de datos masivos	
	RA1	Identificar y organizar temporalmente las tareas necesarias para la realización sus actividades de aprendizaje, cumpliendo con los plazos establecidos de entrega de actividades.
	RA2	Ser capaz de integrarse y participar en el desarrollo organizado de un trabajo en grupo.
CG02	Capacidad de análisis de datos masivos procedentes de diversas fuentes: texto, audio, numérica e imagen	
	RA1 Describir, relacionar e interpretar situaciones y planteamientos sencillos	
	RA2	Seleccionar las fuentes de información digital más significativas
RA3 Identificar las fuentes de información más relevantes para cada problem.		ldentificar las fuentes de información más relevantes para cada problema
	RA4	Utilizar las herramientas digitales idóneas para la adquisición y gestión de la información y es capaz de integrar esa información de una manera raciona
CG03	Resolución de problemas y toma de decisiones en un entorno de datos masivos tanto cuantitativos como cualitativos	
	RA1	Resolver problemas y realizar trabajos aplicando una diversidad metodológica de análisis
	RA2	Identificar correctamente los conocimientos aplicables a cada situación
	RA3	Identificar la metodología más apropiada para la resolución de los problemas planteados, conociendo las herramientas software más relevantes y aportando una visión crítica.
CG07	Capacidad de liderazgo y trabajo en equipo, en la sociedad de la información	
	RA1 Participar de forma activa en el trabajo de grupo compartiendo información, conocimientos y	
	RA2	Orientarse a la consecución de acuerdos y objetivos comunes
	RA3	Ser capaz de organizar y/o coordinar equipos de trabajo
CG08	Capacidad crí	tica y autocrítica en la sociedad de la información
	RA1	Identificar los supuestos y las limitaciones de los modelos y metodologías a aplicar en cada situación
	RA2	Ejercer una valoración crítica de los modelos y metodologías a aplicar en cada situación
CG09	Compromiso ético en la sociedad de la información	
	RA1	Ser honesto en el desarrollo de las actividades académicas y en otros aspectos de la vida y no ser pasivo ante la deshonestidad de otros.



	RA2	Conocer y desarrollar las bases de la ética profesional	
CG11	Capacidad para aprender y trabajar autónomamente en la sociedad de la información		
	RA1 Organizar los materiales de la asignatura con el fin de articular su aprendizaje		
	Poseer criterio para decidir la utilización de los medios humanos y otros recursos puestos a s para el aprendizaje		
	RA3	Realizar sus trabajos y su actividad necesitando sólo unas indicaciones iniciales y un seguimiento básico	
ESPECÍFICAS			
CE25	Comprender los conceptos más importantes asociados al Big Data y sus elementos 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.	
CEO34	Entender la computación y las bases de datos distribuidas, con énfasis en la arquitectura Hadoop y las herramientas actuales más extendidas en el contexto de Big Data.		
	RA01	Conoce y comprende la problemática del Big Data y las técnicas más avanzadas de computación y almacenamiento usadas en la actualidad.	
RA02		Comprende y es capaz de administrar una arquitectura Hadoop y otras arquitecturas NoSQL.	
	RA03	Comprende y aplica las técnicas de MapReduce y otras técnicas avanzadas de computación específicas para el análisis de datos voluminosos.	

BLOQUES TEMÁTICOS Y CONTENIDOS

Contenidos – Bloques Temáticos

- 1. Introducción
- 2. Sistemas Distribuidos
- 3. Linux en Big Data (clase 4)
- 4. Hadoop y presentación
- 5. Almacenamiento HDFS
- 6. Procesamiento Masivo MapReduce, Yarn y Kafka
- 7. Cloud y Al en Big Data



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

HORAS PRESENCIALES				
Lecciones de Carácter expositivo	Exposición pública de temas o trabajos	Ejercicios y resolución de casos y de problemas		
24.00	2.00	6.00		
HORAS NO PRESENCIALES				
Estudios individual y/o en grupo, y lectura organizada	Ejercicios y resolución de casos y de problemas	Trabajos monográficos y de investigación, individuales o colectivos		
30.00	18.00	10.00		
		CRÉDITOS ECTS: 3,0 (90,00 horas)		

EVALUACIÓN Y CRITERIOS DE CALIFICACIÓN

El uso de IA para crear trabajos completos o partes relevantes, sin citar la fuente o la herramienta o sin estar permitido expresamente en la descripción del trabajo, será considerado plagio y regulado conforme al Reglamento General de la Universidad.



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 práctico continuo:	Realización de prácticas de la asignatura	30 %
Evaluación del rendimiento teórico 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

En la convocatoria ordinaria:

- 60% exámenes
 - o Intersemestral: 20%
 - o Final: 40%
- 15% práctica
- 15% práctica
- 10% presentación de grupo

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

- 60% examen extraordinario
- 15% práctica
- 15% práctica
- 10% presentación de grupo

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 contendidos 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, OReilly 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.".
- M. van Steen and A.S. Tanenbaum, Distributed Systems, 3rd ed., distributed-systems.net, 2017.
- Noah Iliinsky ad Julie (2011) Steele Data Visualizations O'Reilly
- Nathan Yau (2011) Visualize This: The Flowing Data 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.".

En cumplimiento de la normativa vigente en materia de **protección de datos de carácter personal**, le informamos y recordamos que puede consultar los aspectos relativos a privacidad y protección de datos <u>que ha aceptado en su matrícula</u> entrando en esta web y pulsando "descargar"

 $\underline{https://servicios.upcomillas.es/sedeelectronica/inicio.aspx?csv=02E4557CAA66F4A81663AD10CED66792}$