



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
Impartido en	Grado en Ingeniería en Tecnologías de Telecom. y Grado en Análisis de Negocios/Business Analytics [Cuarto Curso]
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	Juan Felipe Cerezo Pérez
Departamento / Área	Departamento de Telemática y Computación
Correo electrónico	jfcerezo@icai.comillas.edu

DATOS ESPECÍFICOS DE LA ASIGNATURA

Contextualización de la asignatura
Aportación al perfil profesional de la titulación
<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
Conocimientos básicos de programación.



Competencias - Objetivos

Competencias

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 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 racional
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 experiencias
	RA2	Orientarse a la consecución de acuerdos y objetivos comunes
	RA3	Ser capaz de organizar y/o coordinar equipos de trabajo
CG08	Capacidad crítica 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
	RA2	Poseer criterio para decidir la utilización de los medios humanos y otros recursos puestos a su disposición 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 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.

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.
3. Casos de uso de tecnología 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**



- 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**
- 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

- 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**
- 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**
- 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

Actividades de evaluación	Criterios de evaluación	Peso
<u>Realización de exámenes:</u>	<ul style="list-style-type: none"> Examen final. Comprensión de conceptos mediante pruebas abiertas, enfocado a casos prácticos. Aplicación de los conceptos adquiridos Presentación y comunicación escrita. 	60
<u>Evaluación del rendimiento teórico continuo:</u>	<ul style="list-style-type: none"> 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. 	30



	<ul style="list-style-type: none">• Análisis e interpretación de los resultados obtenidos en la resolución de problemas mediante debates	
<u>Evaluación del rendimiento práctico continuo:</u>	<ul style="list-style-type: none">• 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
<p>Actividades Presenciales y No presenciales</p> <p>Fecha de realización</p> <p>Fecha de entrega</p> <ul style="list-style-type: none">Lectura y estudio de los contenidos teóricos Después de cada claseResolución de los problemas propuestos o debates SemanalmentePreparación de las prácticas Semanas 8, 9 y 10 Semanas 11Preparación 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.



COMILLAS

UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

GUÍA DOCENTE

2023 - 2024

- 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 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.".

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 que ha aceptado en su matrícula entrando en esta web y pulsando "descargar"

[https://servicios.upcomillas.es/sedelectronica/inicio.aspx?csv=02E4557CAA66F4A81663AD10CED66792](https://servicios.upcomillas.es/sedeelectronica/inicio.aspx?csv=02E4557CAA66F4A81663AD10CED66792)