



FICHA TÉCNICA DE LA ASIGNATURA

Datos de la asignatura	
Nombre completo	Tecnologías de procesamiento Big Data
Código	DTC-IMAT-325
Título	Grado en Ingeniería Matemática e Inteligencia Artificial
Impartido en	Grado en Ingeniería Matemática e Inteligencia Artificial [Tercer Curso]
Créditos	6,0 ECTS
Carácter	Obligatoria (Grado)
Departamento / Área	Departamento de Telemática y Computación

Datos del profesorado	
Profesor	
Nombre	Guillermo Gallego Reina
Departamento / Área	Departamento de Telemática y Computación
Correo electrónico	ggallego@icai.comillas.edu
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
Enriquece el perfil profesional al brindar habilidades clave para gestionar eficazmente grandes conjuntos de datos, implementar sistemas distribuidos, automatizar procesos con agentes inteligentes, visualizar datos de manera efectiva, y mantenerse actualizado con las últimas tendencias tecnológicas. Estas competencias son fundamentales para destacar en entornos laborales impulsados por la era del Big Data y la rápida evolución tecnológica.
Prerrequisitos
Debe conocer el funcionamiento de los sistemas distribuidos y el uso de entornos virtualizados. Debe conocer las técnicas de programación adquiridas en cursos anteriores.

Competencias - Objetivos
Competencias
GENERALES



CG04	Conocimientos básicos sobre el uso y programación de los ordenadores, sistemas operativos, bases de datos y programas informáticos con aplicación en ingeniería.
CG05	Conocimiento de la estructura, organización, funcionamiento e interconexión de los sistemas informáticos, los fundamentos de su programación, y su aplicación para la resolución de problemas propios de la ingeniería
CG07	Capacidad para integrarse en equipos de trabajo y colaborar de forma activa con otras personas, áreas y organizaciones en la consecución de los objetivos ligados a las actividades de extracción de valor de los datos e inteligencia artificial.
CG09	Capacidad para determinar eficazmente los objetivos, prioridades, métodos y controles para desempeñar tareas relacionadas con la planificación de proyectos de explotación de datos e inteligencia artificial, mediante la organización de las actividades con los plazos y los medios disponibles

ESPECÍFICAS

CE14	Dominio de los conceptos y técnicas más utilizadas de adquisición y transformación de la información localizada en local o en remoto en el ámbito del análisis de datos y la inteligencia artificial
CE15	Capacidad para diseñar y gestionar sistemas de almacenamiento de información estructurado, semi-estructurado y no estructurado para el desarrollo de aplicaciones en el ámbito de la inteligencia artificial
CE20	Conocimiento de la infraestructura Big Data de almacenamiento y procesamiento distribuido para el procesamiento de datos masivos
CE21	Capacidad para diseñar e implementar aplicaciones Big Data, siendo capaz de identificar y desplegar las tecnologías que mejor se adapten a cada caso de uso para el procesamiento masivo de datos.
CE23	Capacidad para desarrollar y utilizar herramientas de visualización de grandes volúmenes de datos para poder comunicar los resultados de los análisis realizados sobre los mismos.

Resultados de Aprendizaje

RA1	Conocer las etapas de un desarrollo de software Big Data así como los perfiles profesionales que intervienen
RA2	Conocer la base de datos noSQL más extendida en el ecosistema Hadoop
RA3	Desarrollar soluciones de ingesta de datos mediante las herramientas más utilizadas en soluciones Big Data
RA4	Conocer los interfaces de consulta pseudo-SQL para entornos Big Data
RA5	Dominar las tecnologías de mensajería de intercambio de información en streaming en soluciones integrales Big Data
RA6	Conocer, diseñar e implementar aplicaciones utilizando el motor de procesamiento más utilizado para plataformas distribuidas: Spark
RA7	Poseer una visión amplia de las tecnologías existentes (tradicionales e innovadoras) en el ecosistema Big Data

BLOQUES TEMÁTICOS Y CONTENIDOS



Contenidos – Bloques Temáticos

1. Almacenamiento Pseudo-SQL

- Avro.
- Parquet.
- Raw.
- JSON.
- Real Time Data.
- Otros.

2. Ingesta de Datos BigData

- Apache Nifi.
- Apache Sqoop.

3. Sistemas de Mensajes Distribuidos

- Apache Kafka

4. Sistemas de Almacenamiento

- Hbase.
- Hive.
- ElasticSearch.

5. Agentes de Datos

- Beats.
- FileBeat.
- MetricBeat.

6. Visualizador Datos

- Kibanna

7. Motores de Procesamiento de Big Data

- SPARK
- Arquitectura.
- Diseño Aplicaciones.
- Rendimiento.
- Visualización de Datos.
- Streaming: SCALA.
- MLIB.
- GRAPHX.

8. Productividad Modelos

- Flask RestFul Api

9. Tendencias Tecnológicas

- IA Generativa.

METODOLOGÍA DOCENTE

Aspectos metodológicos generales de la asignatura

Metodología Presencial: Actividades

Las actividades formativas se desarrollarán durante las 4 horas de clase a la semana que se distribuirán:



- **Clases magistrales expositivas y participativas:**
 - El profesor realizará un exposición de los contenidos teóricos.
- **Ejercicios prácticos y resolución de problemas:**
 - El alumno planteará dudas sobre los conceptos teóricos expuestos en la clase magistral, y de las prácticas propuestas.
- **Sesiones prácticas con uso de software:**
 - Se dedicará tiempo al entendimiento y resolución de dichas sesiones.
- **Actividades de evaluación continua del rendimiento:**
 - Se realizarán pruebas, desarrollarán prácticas complementarias a las semanales y retos gamificados.
- **Tutoría para resolución de dudas:**
 - Se realizará de forma implícita durante el resto de actividades descritas.

CG04, CG05, CG07, CG09, CE14, CE15, CE20, CE21, CE23

Metodología No presencial: Actividades

Las actividades formativas serán:

- **Ejercicios prácticos y resolución de problemas:**
 - El alumno dispondrá de problemas concretos enfocados a asimilar los conceptos explicados en las sesiones teóricas para desarrollar de forma no presencial. La solución de estos problemas será subida a la plataforma.
- **Sesiones prácticas con uso de software:**
 - Una vez liberada la práctica, el alumno trabajará sobre ella de forma no presencial.
- **Estudio personal:**
 - El objetivo principal del trabajo no presencial es llegar a entender y comprender los conceptos teóricos de la asignatura, así como ser capaz de poner en práctica estos conocimientos para resolver los diferentes tipos de problemas.

CG04, CG05, CG07, CG09, CE14, CE15, CE20, CE21, CE23

RESUMEN HORAS DE TRABAJO DEL ALUMNO

HORAS PRESENCIALES			
Clases magistrales expositivas y participativas	Sesiones prácticas con uso de software	Actividades de evaluación continua del rendimiento	Ejercicios prácticos y resolución de problemas
40.00	15.00	2.00	2.00
HORAS NO PRESENCIALES			
Ejercicios prácticos y resolución de problemas	Sesiones prácticas con uso de software	Estudio personal	Proyectos
3.00	32.00	30.00	50.00
CRÉDITOS ECTS: 6,0 (174,00 horas)			

EVALUACIÓN Y CRITERIOS DE CALIFICACIÓN

Actividades de evaluación	Criterios de evaluación	Peso
Examen:	<ul style="list-style-type: none"> • Examen Final (50%): <ul style="list-style-type: none"> ◦ Se evaluará los conocimientos 	50 %



<ul style="list-style-type: none">Examen Final.	adquiridos en la asignatura.	
Sesiones prácticas: <ul style="list-style-type: none">Retos Colaborativos.Trabajos no presenciales.Entregables	La actitud, participación y realización de los entregables y los retos planteados en sesiones colaborativas e individuales.	20 %
Proyecto final	Proyecto final de la asignatura que el alumno entregará al finalizar el curso.	30 %

Calificaciones

La calificación final en convocatoria ordinaria y extraordinaria de la asignatura dependerá de la evaluación de las siguientes actividades:

Nota Final = 50% Examen_Final + 20% Entregables + 30% Proyecto final

Para aprobar la asignatura los alumnos tienen que obtener al menos 5 puntos sobre 10 en el examen final de la asignatura y en la práctica final, tanto en la convocatoria ordinaria como en la extraordinaria. La inasistencia al 15% o más de las horas presenciales de esta asignatura puede tener como consecuencia la imposibilidad de presentarse a las convocatorias ordinaria y extraordinaria.

BIBLIOGRAFÍA Y RECURSOS

Bibliografía Complementaria

1. Transactional Machine Learning with Data Streams and AutoML: Build Frictionless and Elastic Machine Learning Solutions with Apache Kafka in the Cloud Using Python.
2. Learn Apache NiFi: Data Integration and Flow Orchestration: A Comprehensive Guide to Designing, Managing, and Optimizing Data Workflows.
3. Designing Machine Learning Systems: An Iterative Process for Production-Ready Applications. O'Reilly.
4. Learning Spark: Lightning-fast Data Analytics. O'Reilly.
5. Practical MLOps: Operationalizing Machine Learning Models.

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>