



FICHA TÉCNICA DE LA ASIGNATURA

Datos de la asignatura

Nombre completo	Optativa Complementaria. Procesamiento en Streaming
Código	DTC-MBD-525
Título	Máster Universitario en Big Data por la Universidad Pontificia Comillas
Nivel	Posgrado Oficial Master
Cuatrimestre	Semestral
Créditos	3,0 ECTS
Carácter	Obligatoria
Departamento / Área	Departamento de Telemática y Computación
Responsable	Patricia Alcalde

Datos del profesorado

DATOS ESPECÍFICOS DE LA ASIGNATURA

Contextualización de la asignatura

Aportación al perfil profesional de la titulación

El propósito del curso es que los alumnos adquieran conocimientos teóricos y principalmente prácticos de cómo analizar y procesar datos en tiempo real y qué herramientas son las más adecuadas para hacerlo dentro del ecosistema bin data.

Al final del curso se espera que los alumnos adquieran las siguientes habilidades:

- Conocimientos de los principios básicos del procesamiento en streaming.
- Experiencia práctica con las herramientas Spark Streaming y Kafka principalmente.
- Programación en scala orientada a integrarlo con Spark streaming.
- Conocimiento de otras herramientas del mercado como kafka streams o flink.

Prerrequisitos

- Es aconsejable que los alumnos tengan conocimientos de algún lenguaje de programación como Java , aunque no necesariamente orientado a objetos.
- También es deseable conocimientos básicos de herramientas de bus de datos, como Kafka/PubSub, etc.

Competencias - Objetivos

Competencias

Conocimientos o contenidos



CO1	Entender los fundamentos de la analítica de datos y su aplicación en diversas áreas de la inteligencia artificial, destacando la integración en soluciones complejas y multidisciplinares para el análisis avanzado de datos masivos atendiendo a la diversidad de problemas específicos de cada área.
CO2	Comprender las técnicas de procesamiento de datos, las arquitecturas y herramientas más habituales y apropiadas para condiciones y requisitos de casos específicos.
Competencias	
CP1	Integrar las arquitecturas, técnicas de inteligencia artificial, análisis avanzado de datos y de visualización y de cumplimiento legal para ofrecer la solución global óptima.
CP2	Aplicar e integrar los flujos programáticos de datos masivos
CP4	Implementar las técnicas de procesamiento de datos y usar las herramientas más habituales y apropiadas a las condiciones y requisitos de casos específicos.
CP7	Aplicar conocimientos avanzados en Big Data y analítica de datos para desarrollar soluciones innovadoras en proyectos y en investigación, aportando y evaluando soluciones óptimas para el procesamiento y análisis de datos a gran escala.
Habilidades o destrezas	
HA1	Comunicar de manera oral y escrita con rigor técnico, claridad expositiva y coherencia argumentativa a todo tipo de interlocutores, técnicos y no técnicos.
HA2	Trabajar en equipos de carácter pluridisciplinar y/o internacional y organizar y liderar adecuadamente las dinámicas de grupo.
HA3	Desarrollar las habilidades interpersonales que requieren los entornos profesionales actuales (empatía, tolerancia, respeto, capacidad para aunar intereses contrapuestos).
HA4	Gestionar, organizar y planificar adecuadamente el trabajo y el tiempo, cumpliendo objetivos y estándares de calidad.
HA5	Mantener una formación y aprendizaje continuo y adaptación a los cambios tecnológicos y científicos.

BLOQUES TEMÁTICOS Y CONTENIDOS

Contenidos – Bloques Temáticos

1. Sistemas de mensajería distribuidos con colas de eventos.
2. Tratamiento de eventos en streaming con una herramienta adecuada para grandes cantidades de datos.
3. Lenguaje de programación para procesamiento de datos en streaming.
4. Disponibilización de datos en real time para una posterior explotación de los mismos.
5. Iniciación a tecnologías devops.



6. Overview de tecnologías de procesamiento de datos en streaming.

METODOLOGÍA DOCENTE

Aspectos metodológicos generales de la asignatura
Metodología Presencial: Actividades
Clases teóricas con los fundamentos básicos del curso , principalmente focalizados en conceptos de streaming y el procesamiento de datos. Ejercicios guiados y resueltos por el profesor en clase de cada uno de los módulos más relevantes del curso.
Metodología No presencial: Actividades
Prácticas de realización individual y evaluables sobre los contenidos de la asignatura.

RESUMEN HORAS DE TRABAJO DEL ALUMNO

HORAS PRESENCIALES	
Clases magistrales expositivas y participativas: Exposición de contenidos fundamentales por parte del profesor impulsando la reflexión y participación del alumno.	Ejercicios prácticos y resolución de problemas: Sesiones prácticas con uso de software: Actividad formativa con ordenador que, bajo la guía del profesor-tutor, fomenta el aprendizaje autónomo y/o cooperativo del alumno mediante la ejecución de programas para la consecución de los objetivos marcados
15.00	15.00
HORAS NO PRESENCIALES	
Estudio personal: Reflexión y análisis individual de los contenidos teóricos y prácticos de las materias y/o asignaturas del Master	Trabajos: Los alumnos tendrán que hacer trabajos breves (individuales y/o en grupo), por indicación del profesor
45.00	15.00
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
---------------------------	-------------------------	------



Durante el curso se pedirá realizar a los alumnos prácticas individuales para asentar los conocimientos de la asignatura y hacerles un seguimiento continuo.	Se valorará que las funciones estén correctamente implementadas , que el Outputs sea correcto, que se hayan implementado tests y que sea programación funcional. Se debe tener en media , un 5 en este bloque de prácticas para superar la asignatura.	60
El examen tendrá un parte de test y otra de práctica.	Se debe tener en media , un 5 en el examen para superar la asignatura.	40

BIBLIOGRAFÍA Y RECURSOS

Bibliografía Básica

Durante el curso seguiremos la documentación oficial de Apache Spark:

<https://spark.apache.org/docs/2.4.0/structured-streaming-programming-guide.html>

Bibliografía Complementaria

Como bibliografía complementaria se sugiere: Programming Scala O'Reilly