



FICHA TÉCNICA DE LA ASIGNATURA

Datos de la asignatura	
Nombre completo	Arquitectura Big Data
Código	DTC-MBD-511
Título	Máster Universitario en Big Data por la Universidad Pontificia Comillas
Impartido en	Máster Universitario en Big Data [Primer Curso]
Créditos	3,0 ECTS
Carácter	Obligatoria
Departamento / Área	Departamento de Telemática y Computación

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>Capacidad de entender los principios de sistemas distribuidos y su aplicación en el entorno de Big Data.</p> <p>Experiencia práctica algunas de las principales herramientas empleadas en el desarrollo de proyectos Big Data.</p> <p>Aplicación de los conceptos de sistemas distribuidos en las herramientas Big Data más utilizadas.</p> <p>Conceptos básicos del funcionamiento de la virtualización</p>	
Prerrequisitos	
<p>Conocimientos básicos de redes TCP/IP, sistemas operativos y arquitectura de ordenadores.</p> <p>Conocimientos básicos de programación y shell scripting de unix.</p>	

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 procesados 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.
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

Bloques Temáticos

Bloque 1: Conceptos Arquitecturales, Metodologías de trabajo. Conceptos de sistemas distribuidos y su aplicación a Big Data

Bloque 2: Principales herramientas de Big Data: HDFS, YARN y Kafka. Paradigma Map-Reduce

Bloque 3: Virtualización, conceptos básicos de contenedores y máquinas virtuales. Entornos de red y disco en virtualización

Contenidos

Clase 1 : Introducción a la Arquitectura Big Data

Clase 2 : Conceptos de Sistemas Distribuidos I

Clase 3 : Conceptos de Sistemas Distribuidos II

Clase 4 : HDFS



Clase 5 : Kafka
 Clase 6 : Map-Reduce / Yarn
 Clase 7 : Otras herramientas no Hadoop
 Clase 8 : Virtualizacion I
 Clase 9 : Virtualizacion II
 Clase 10 : Practica con herramientas en Cluster
 Clase 11 : Dimensionamiento
 Clase 12 : Practica de caso de Uso
 Clase 13 : Arquitecturas de BI y Metodologias
 Clase 14 : Presentación de herramientas Big Data

METODOLOGÍA DOCENTE

Aspectos metodológicos generales de la asignatura

La metodología a seguir será la siguiente:

- Clases magistrales donde el profesor exponga los principales elementos
- Ejercicios guiados en clase donde los alumnos vayan explorando de manera individual pero guiada los elementos mas relevantes de la asignatura
- Ejercicios individuales de los alumnos para que desarrollen en mas profundidad y de manera personal los contenidos de la asignatura
- Proyectos de trabajo en grupo para que los alumnos de manera colaborativa trabajen un caso de uso y el estudio de una herramienta determinada, así como su presentación al resto de la clase

Metodología Presencial: Actividades

Clases teórico-prácticas: Presentación de los conceptos básicos por parte del profesor. Uso, configuración o implementación de estos conceptos empleando diferentes herramientas dentro del cluster de servidores de Big Data

CO1, CO2, CP1, CP7

Hackaton: Los alumnos tendrán que aplicar los conceptos de sistemas distribuidos en una competición que constará de diferentes pruebas prácticas. Estas pruebas se realizarán por grupos que se intentará que tengan un nivel lo más homogéneo posible. Cada grupo tendrá que definir su estrategia de trabajo: el orden en que afrontarán las pruebas y cuantos miembros del equipo las abordarán. Los grupos de alumnos tendrán que trabajar de manera autónoma, pero también tendrán la posibilidad de solicitar ayuda al profesor en un número limitado de ocasiones.

CO2, CP1, HA2, HA3,
HA4, HA5

Presentación de una herramienta Big Data: En grupos, los alumnos tendrán que estudiar alguna de las herramientas Big Data mas empleadas en la industria y hacer una presentación de la misma al resto de la clase. Tendrán que hacer foco en los conceptos principales de sistemas distribuidos: replicación, paralelismo y alta disponibilidad.

CO1, CO2, HA1, HA3

Caso de uso de una arquitectura Big Data: en grupos los alumnos tendrán que desarrollar una arquitectura de Big Data para un caso de negocio determinado. Tendrán que descomponer la arquitectura en sus diferentes elementos y, aplicando los conocimientos teóricos, así como los conocimientos espécificos ya adquiridos sobre las diferentes herramientas, proponer una arquitectura que pueda presentar una solución al

CO1, CP4, CP7, HA2,
HA3



problema presentado	
Metodología No presencial: Actividades	
Estudio personal	CO1, CO2, CP1, CP4, CP7
Trabajo en equipo para la preparación de la presentación de la herramienta de Big Data	
Trabajo y desarrollo de las ejercicios de presentación obligatoria	CO2, CP1, CP7, HA2, HA3, HA4

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
15.00	30.00
CRÉDITOS ECTS: 3,0 (75,00 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
Examen escrito. Consistirá en preguntas de respuesta breve donde los alumnos tendrá que conocer los contenidos de la asignatura y aplicarlos a algunos casos prácticos sencillos. También incluirá un problema de dimensionamiento hardware.	Demostrar en cada una de las preguntas que se conocen y aplican los conocimientos teóricos de la asignatura de manera correcta. Las respuestas deben ser breves y concretas, centradas en contestar a la pregunta planteada, no a otros elementos que, aunque sean de interés, no corresponden a la pregunta realizada.	65
	Se valorarán los siguientes elementos:	



<p>Prácticas individuales.</p> <p>Prácticas de grupo</p>	<ul style="list-style-type: none"> • una correcta presentación de los resultados en los ejercicios escritos • que los resultados se ajusten adecuadamente a las preguntas realizadas • dominio y soltura en el manejo de los conceptos en las presentaciones orales • correcto funcionamiento de las herramientas en las pruebas del Hackaton 	35
--	---	----

Calificaciones

Para superar la asignatura la puntuación del examen debe ser mayor que 5 sobre 10.

Para superar la asignatura, la puntuación de la media de las prácticas debe ser mayor que 5 sobre 10.

La asistencia al 85% de las clases es obligatoria para poder hacer el examen, según el Artículo 93 del Reglamento General de la Universidad de Comillas.

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

PLAN DE TRABAJO Y CRONOGRAMA

Actividades	Fecha de realización	Fecha de entrega
Clases 1 y 2 - Introducción a la Arquitectura Big Data, Arquitecturas de BI y Metodologías	Semana 1	
Clases 3 y 4 : Conceptos de Sistemas Distribuidos	Semana 2	
Clases 5 a 8 : Herramientas Big Data, HDFS, Map-Reduce / Yarn, Kafka	Semana 3 - Semana 4	
Clases 9 y 10 : Virtualización	Semana 5	
Clase 11 - Hackaton	Semana 6	
Clase 12 - Dimensionamiento del hardware de sistemas Big Data	Semana 6	
Práctica de desarrollo de un procesamiento Big Data		Semana 4
Práctica de funcionamiento sobre HDFS		Semana 5
Presentación de una herramienta de Big Data		Semana 7
Diseño de una arquitectura de Big Data empleando las herramientas conocidas y ejemplos de la industria		Semana 7



BIBLIOGRAFÍA Y RECURSOS

Bibliografía Básica

Presentaciones teóricas de la asignatura

Documentación oficial de las principales herramientas presentadas: Kafka, HDFS, YARN, Map-Reduce

Bibliografía Complementaria

Hadoop: The Definitive Guide, 4th Edition, ISBN:9781491901687

Kafka: The Definitive Guide, 2nd Edition, ISBN: 9781492043072

Designing Distributed Systems, ISBN: 9781491983638