

Departamento de Informática

Curso de Especialización

Inteligencia Artificial y Big Data

Programación didáctica del Módulo:

Sistemas Big Data

Curso 2023-2024

**Profesores: Luis García Bonifaz
Ricardo Álvaro Blanco**

1.- PROPUESTAS DE MEJORA DEL CURSO ANTERIOR3

2.- OBJETIVOS DEL MÓDULO	3
3.- CONTENIDOS.....	6
4.- DISTRIBUCIÓN TEMPORAL.....	6
5.- METODOLOGÍA DIDÁCTICA	7
6.- RESULTADOS DE APRENDIZAJE Y CRITERIOS DE EVALUACIÓN	8
7.- PROCEDIMIENTOS Y INSTRUMENTOS DE CUALIFICACIÓN	10
8.- MATERIALES Y RECURSOS DIDÁCTICOS	13
9.- ACTIVIDADES COMPLEMENTARIAS I EXTRAESCOLARES	14
10.- TEMAS TRANSVERSALES.....	14
11.- ATENCIÓN A LA DIVERSIDAD	14

1.- PROPUESTAS DE MEJORA DEL CURSO ANTERIOR

El curso pasado se propusieron las siguientes mejoras:

- Reducir el número de horas dedicadas al Ecosistema Hadoop
- Reducir el número de horas dedicadas a Spark RDD y ampliar las de Spark SQL y Spark Streaming
- Incluir uno o más proyectos que se desarrollen a lo largo del curso

Todas estas propuestas han sido incluidas en la presenta programación.

Se incluye además el estudio de Kafka para tratamiento de datos en tiempo real.

2.- OBJETIVOS DEL MÓDULO

Como viene especificado en el Artículo 4 del Real Decreto 279/2021, de 20 de abril, por el que se establece el Curso de especialización en Inteligencia Artificial y Big Data y se fijan los aspectos básicos del currículo, la competencia general en la que se enmarca este módulo *“consiste en programar y aplicar sistemas inteligentes que optimizan la gestión de la información y la explotación de datos masivos, garantizando el acceso a los datos de forma segura y cumpliendo los criterios de accesibilidad, usabilidad y calidad exigidas en los estándares establecidos, así como los principios éticos y legales”*.

En concreto en este módulo las funciones principales que debe conocer el alumnado son:

- Tener una visión global de lo que es Big Data
- Conocer las bases de datos NoSQL, en especial MongoDB. Ser capaz de crear, consultar y modificar datos en esta base de datos documental.
- Conocer el ecosistema de ficheros distribuidos Hadoop
- Utilizar la herramienta Spark par consultar datos de Hadoop tanto en procesos batch como en streaming
- Conocer a grandes rasgos los servicios que ofrecen las grandes empresas de almacenamiento en la nube como son Amazon Web Services y Microsoft Azure

Y con ello poder alcanzar los siguientes objetivos generales:

- a) Aplicar técnicas de tratamiento de datos para gestionar la transformación digital en las organizaciones.
- b) Utilizar soluciones de Big Data para integrar sistemas de explotación de datos.
- c) Analizar y evaluar soluciones Big Data para su implantación en las funcionalidades, procesos y sistemas de decisiones.
- d) Determinar la documentación técnica y normativa vigente de los procedimientos de protección de datos para ejecutar el sistema de explotación de datos cumpliendo con los principios legales y éticos.
- e) Determinar la solución de Inteligencia Artificial y Big Data para configurar las herramientas y lenguajes específicos.
- f) Aplicar técnicas Big Data para gestionar los datos de la organización y obtener conocimiento a partir de ellos.
- g) Analizar y utilizar los recursos y oportunidades de aprendizaje relacionados con la evolución científica, tecnológica y organizativa del sector y las tecnologías de la información y la comunicación, para mantener el espíritu de actualización y adaptarse a nuevas situaciones laborales y personales.
- h) Desarrollar la creatividad y el espíritu de innovación para responder a los retos que se presentan en los procesos y en la organización del trabajo y de la vida personal.
- i) Evaluar situaciones de prevención de riesgos laborales y de protección ambiental, proponiendo y aplicando medidas de prevención personal y colectiva, de acuerdo con la normativa aplicable en los procesos de trabajo, para garantizar entornos seguros.
- j) Identificar y proponer las acciones profesionales necesarias, para dar respuesta a la accesibilidad universal, al «diseño para todas las personas», así como para evitar posibles sesgos de género en el desarrollo y aplicaciones de Inteligencia Artificial y Big Data.
- k) Identificar y aplicar parámetros de calidad en los trabajos y actividades realizados en el proceso de aprendizaje, para valorar la cultura de la evaluación y de la calidad y ser capaces de supervisar y mejorar procedimientos de gestión de calidad.

Y las siguientes competencias profesionales y sociales:

- a) Administrar el desarrollo de procesos automatizados que permitan la mejora de la productividad de las empresas.
- b) Integrar sistemas de explotación de grandes volúmenes de datos aplicando soluciones de Big Data.
- c) Ejecutar el sistema de explotación de datos según las necesidades de uso y las condiciones de seguridad establecidas asegurando el cumplimiento de los principios legales y éticos.
- d) Gestionar de manera eficiente los datos, la información y su representación para transformarlos en conocimiento.
- e) Cumplir la legislación vigente que regula la normativa de los medios de comunicación audiovisual y de la accesibilidad universal.
- f) Adaptarse a las nuevas situaciones laborales, manteniendo actualizados los conocimientos científicos, técnicos y tecnológicos relativos a su entorno profesional, gestionando su formación y los recursos existentes en el aprendizaje a lo largo de la vida y utilizando las tecnologías de la información y la comunicación.
- g) Resolver situaciones, problemas o contingencias con iniciativa y autonomía en el ámbito de su competencia, con creatividad, innovación y espíritu de mejora en el trabajo personal y en el de los miembros del equipo.
- h) Organizar y coordinar equipos de trabajo con responsabilidad, supervisando el desarrollo del mismo, manteniendo relaciones fluidas y asumiendo el liderazgo, así como aportando soluciones a los conflictos grupales que se presenten.
- i) Comunicarse con sus iguales, superiores, clientela y personas bajo su responsabilidad, utilizando vías eficaces de comunicación, transmitiendo la información o conocimientos adecuados y respetando la autonomía y competencia de las personas que intervienen en el ámbito de su trabajo.
- j) Generar entornos seguros en el desarrollo de su trabajo y el de su equipo, supervisando y aplicando los procedimientos de prevención de riesgos laborales y ambientales, de acuerdo con lo establecido por la normativa y los objetivos de la empresa.
- k) Supervisar y aplicar procedimientos de gestión de calidad, de accesibilidad universal y de «diseño para todas las personas», en las actividades profesionales incluidas en los procesos de producción o prestación de servicios.

- l) Ejercer sus derechos y cumplir con las obligaciones derivadas de su actividad profesional, de acuerdo con lo establecido en la legislación vigente, participando activamente en la vida económica, social y cultural.

3.- CONTENIDOS

T1.- Introducción Big Data

T2.- Big Data con (AWS)

T3.- Servicios AWS para Big Data

T4.- Ecosistema Hadoop

- Map Reduce
- YARN + Hive + Impala

T5.- Clusters

T6.- Procesamiento en tiempo real.

- Kafka

T7.- Spark

- Spark RDD
- Spark SQL
- Spark Streaming

4.- DISTRIBUCIÓN TEMPORAL

La distribución temporal de los contenidos será la mostrada en la siguiente tabla:

Tema	Contenido	Horas	Evaluación
T1	Introducción Big Data	2	1

T2	Big Data Con AWS	2	1
T3	Servicios AWS para Big Data	42	1 y 2
T3	Ecosistema Hadoop	4	1
T4	Clusters	4	1
T5	Procesamiento en tiempo real.	6	1
T6	Spark	36	2
Total ...		90	

El reparto de los temas se hará en dos bloques que se impartirán en paralelo:

- Un bloque con una carga lectiva de 1 hora semanal en el que se verá el tema 2.
- Otro bloque de 2 horas semanales en el que se verán el resto de los temas.

5.- METODOLOGÍA DIDÁCTICA

La metodología que se utilizará en el módulo busca potenciar los siguientes aspectos:

- Utilización de todos los recursos que brindan las nuevas tecnologías.
- Aprendizaje autónomo, activo y participativo.
- Evaluación personalizada y continua.

La didáctica del módulo se fundamentará en los siguientes aspectos:

- Aunque se utilizarán las explicaciones del profesor para introducir conceptos y aclarar posibles ideas erróneas, la metodología será principalmente procedimental.
- Las explicaciones se pondrán siempre en práctica a través de ejercicios y práctica con

el ordenador.

- Se propondrán actividades para fomentar el uso de manuales.
- Se fomentará el aprendizaje autónomo de los alumnos mediante la tutorización del profesor.
- Se podrán realizar trabajos de investigación para el desarrollo de contenidos.
- Se realizará al menos un examen teórico/práctico en cada evaluación. Los exámenes podrán constar de preguntas de tipo test, preguntas de respuesta corta, preguntas de desarrollo de contenidos y ejercicios prácticos en los que normalmente será necesario el uso del ordenador o de manuales.
- El profesor organizará las tareas para combinar el trabajo individual y el trabajo en grupo a lo largo del curso.
- Los alumnos podrán realizarán exposición de trabajos con el apoyo de material multimedia.

6.- RESULTADOS DE APRENDIZAJE Y CRITERIOS DE EVALUACIÓN

Relacionados con los niveles de competencia profesional descritos en el currículo del curso y mencionados el punto 2 se citan a continuación los resultados de aprendizaje y criterios de evaluación asignados a este módulo.

1. Aplica técnicas de análisis de datos que integran, procesan y analizan la información, adaptando e implementando sistemas que las utilicen.

Criterios de evaluación:

- a) Se han identificado conceptos básicos de matemática discreta, lógica algorítmica y complejidad computacional, y su aplicación para el tratamiento automático de la información por medio de sistemas computacionales.
- b) Se ha extraído de forma automática información y conocimiento a partir de grandes volúmenes de datos.

- c) Se han combinado diferentes fuentes y tipos de datos.
- d) Se ha construido un conjunto de datos complejos y se han relacionado entre sí.
- e) Se han establecido objetivos y prioridades, secuenciación y organización del tiempo de realización.
- f) Se han seleccionado e integrado sistemas de información que satisfacen las necesidades del problema.
- g) Se han determinado criterios de coste y calidad necesarios para la eficacia y eficiencia de la implementación de un sistema Big Data.

2. Configura cuadros de mando en diferentes entornos computacionales usando técnicas de análisis de datos.

Criterios de evaluación:

- a) Se han clasificado diferentes librerías e implementaciones de las técnicas de representación de la información.
- b) Se ha cruzado información sobre el objetivo a conseguir y la naturaleza de los datos.
- c) Se ha realizado un cuadro de mandos utilizando técnicas sencillas.
- d) Se han utilizado técnicas predictivas complejas para anticiparse a lo que ocurra.
- e) Se ha evaluado el impacto del análisis de datos en la consecución de los objetivos propuestos.

3. Gestiona y almacena datos facilitando la búsqueda de respuestas en grandes conjuntos de datos.

Criterios de evaluación:

- a) Se han extraído y almacenado datos de diversas fuentes, para ser tratados en distintos escenarios.
- b) Se ha fijado el objetivo de extraer valor de los datos para lo que es necesario contar con

tecnologías eficientes.

- c) Se ha comprobado que la revolución digital exige poder almacenar y procesar ingentes cantidades de datos de distinto tipo y descubrir su valor.
- d) Se han desarrollado sistemas de gestión, almacenamiento y procesamiento de grandes volúmenes de datos de manera eficiente y segura, teniendo en cuenta la normativa existente.
- e) Se han utilizado habilidades científicas en entornos de trabajo multidisciplinares.

4. Aplica herramientas para la visualización de datos utilizadas en las soluciones Big Data facilitando las tareas de análisis y presentación de resultados.

Criterios de evaluación:

- a) Se han examinado distintos escenarios y tipologías de datos no estructurados.
- b) Se ha implantado la aplicación de la BI (Business Intelligence) para la extracción de valor.
- c) Se ha reconocido la importancia de almacenar grandes volúmenes de datos de forma distribuida y redundante en un clúster de máquinas.
- d) Se han determinado las diferencias en el entorno de aplicaciones relacionadas que facilitan el procesamiento de datos de manera rápida, eficiente y eficaz.
- e) Se ha comprobado la manera de programar y procesar automáticamente la estructura de datos.
- f) Se han valorado las diferentes formas de visualizar los datos que nos interese representar gráficamente, facilitando así las tareas de análisis y presentación de resultados.

7.- PROCEDIMIENTOS Y INSTRUMENTOS DE CUALIFICACIÓN

La evaluación será continua y se evaluará el trabajo del alumnado a lo largo de todo el curso.

La asistencia a las clases es obligatoria y tal como se recoge en las instrucciones de principio de curso, las faltas de asistencia no justificadas a un 15% o más de las horas del módulo, conllevará

la pérdida de la evaluación continua.

El alumno siempre podrá ser requerido por el profesor para defender y argumentar sobre cualquiera de los trabajos desarrollados y entregados durante el curso escolar.

Las calificaciones en este módulo vendrán dadas por la superación y el dominio de:

- Los contenidos conceptuales (hechos, conceptos y principios), recogidos en las capacidades terminales y criterios de evaluación expuestos anteriormente.
- Se evaluarán a través de:
 - Ejercicios, prácticas y trabajos de investigación.
 - Exámenes.
- Los contenidos procedimentales (“saber cómo hacer” o “saber hacer”) que expresan las habilidades cognitivas (aplicación, análisis, síntesis, evaluación) del alumno.
- Los contenidos actitudinales (“saber ser y estar”) que expresan la auto regulación del comportamiento en función del rol profesional.

El alumno tiene una serie de actividades obligatorias, calificadas de 0 a 10 y además habrá un examen por evaluación calificado también de 0 a 10.

La nota de cada evaluación se calculará con la siguiente formula:

$$\text{Nota} = (\text{Nota Examen} * 0,7) + (\text{Nota Media de las actividades} * 0,3)$$

La nota de cada evaluación se redondeará al alza a la unidad en los boletines de cada evaluación. Conservando los decimales para el cálculo de la nota final.

La nota final se calculará como la media de las 2 evaluaciones.

A continuación, se muestra el cuadro resumen del cálculo de la nota

Aclaraciones:

La falta de autenticidad en la autoría o de originalidad de las pruebas de evaluación; la copia o el plagio; el intento fraudulento de obtener un resultado académico mejor; la colaboración, el encubrimiento o el favorecimiento de la copia, o la utilización de material, aplicaciones o dispositivos no autorizados durante la evaluación, entre otras, son conductas irregulares que pueden tener consecuencias académicas y disciplinarias graves.

Por un lado, si se detecta alguna de estas conductas irregulares, puede comportar el suspenso en las actividades evaluables o en la calificación final de la asignatura.

Por otro lado, y de acuerdo con la normativa académica, las conductas irregulares en la evaluación, además de comportar el suspenso de la asignatura, pueden dar lugar a la incoación de un procedimiento disciplinario y a la aplicación, si procede, de la sanción correspondiente.

	INSTRUMENTOS DE CALIFICACION	%	CONSIDERACIONES
1ª Evaluación	Exámenes	70	Valores aproximados
	Prácticas	30	
2ª Evaluación	Exámenes	70	Valores aproximados
	Prácticas	30	
RECUPERACIONES	No hay recuperaciones		
FINAL ORDINARIA (Junio)	Examen	100	El alumno se examinará de todas aquellas evaluaciones que él tenga suspendidas o no haya podido acudir con causa justificada. La nota final será la nota media de las 2 evaluaciones.
FINAL EXTRAORDINARIA	Examen	100	Será de todo el curso
Alumnado que ha perdido el derecho a la evaluación continua		100	Será de todo el curso

8.- MATERIALES Y RECURSOS DIDÁCTICOS

El material utilizado durante el curso será el siguiente:

Material bibliográfico:

- No se utilizará un libro de texto específico.
- Apuntes del profesor en formato electrónico
- Artículos, vídeos y material multimedia sobre SO.

Materiales físicos:

- 24 ordenadores más uno para el profesorado
- Servidor con Ubuntu Server
- Proyector.
- Pizarra.

Plataforma de enseñanza:

- Campus virtual Aules
- AWS Academy

9.- ACTIVIDADES COMPLEMENTARIAS I EXTRAESCOLARES

No se ha planteado la realización de ninguna actividad complementaria. No obstante, si hubiese alguna opción se planteará a lo largo del curso la realización de estas actividades.

10.- TEMAS TRANSVERSALES

Tal y como aparece en el currículo del curso cada vez cobra más importancia el refuerzo de las competencias de carácter transversal, como el trabajo en equipo, la capacidad de liderazgo, el conocimiento de las tecnologías de la información y comunicación, la visión global, planificación y organización, entre otros.

Los contenidos teóricos, así como las actividades planteadas en clase recogerán este refuerzo de las competencias transversales.

11.- ATENCIÓN A LA DIVERSIDAD

La aplicación en momentos puntuales de técnicas de aprendizaje cooperativo tendrá como una de sus finalidades en tratamiento de la diversidad a través de la confección de grupos heterogéneos y la forma en que se concibe el aprendizaje como una tarea compartida, que sea caza con la ayuda del grupo.

De forma extraordinaria si se observan situaciones de carencia de conocimientos y problemas de aprendizaje se tomarán medidas de refuerzo particulares a través de la asignación de actividades extra.



CIPFP Mislata
Centre Integrat Públic
Formació Professional Superior

PROGRAMACIÓN DIDÁCTICA

