

Gestión de Datos

Código: 106566
Créditos ECTS: 6

2024/2025

Titulación	Tipo	Curso
2504392 Inteligencia Artificial / Artificial Intelligence	OB	3

Contacto

Nombre: Javier Panadero Martinez

Correo electrónico: javier.panadero@uab.cat

Equipo docente

Alvaro Wong Gonzalez

Idiomas de los grupos

Puede consultar esta información al [final](#) del documento.

Prerrequisitos

Aunque no hay prerrequisitos formalmente establecidos y es responsabilidad de la propia asignatura proporcionar a los alumnos un medio para adquirir los conocimientos descritos en el apartado de contenidos de la asignatura. Es recomendable: un buen conocimiento de programación, del funcionamiento de un computador, del sistema operativo a nivel de usuario programador y de los sistemas de bases de datos.

Objetivos y contextualización

El objetivo de esta asignatura es conocer los conceptos fundamentales del diseño de las aplicaciones intensivas de datos respecto a su fiabilidad, escalabilidad y sostenibilidad.

Se estudiarán los sistemas distribuidos y los paradigmas de procesamiento masivo de datos. Se revisarán diversos modelos de programación para aplicaciones de procesamiento "batch", en memoria y "streaming". Se presentarán diversas arquitecturas de aplicaciones comunes y se presentarán conceptos relevantes como integridad, accesibilidad, fiabilidad, consistencia y seguridad en el procesamiento de datos a gran escala.

Competencias

- Analizar y resolver problemas de forma efectiva, generando propuestas innovadoras y creativas para alcanzar los objetivos.

- Conceptualizar y modelar alternativas de soluciones complejas a problemas de aplicación de la inteligencia artificial en diferentes ámbitos, y planificar y gestionar proyectos para el diseño y desarrollo de prototipos que demuestren la validez del sistema propuesto.
- Conocer y utilizar de forma eficiente las técnicas y herramientas de representación, manipulación, análisis y gestión de datos a gran escala.
- Introducir cambios en los métodos y los procesos del ámbito de conocimiento para dar respuestas innovadoras a las necesidades y demandas de la sociedad.
- Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio.
- Trabajar cooperativamente para la consecución de objetivos comunes, asumiendo la propia responsabilidad y respetando el rol de los diferentes miembros del equipo.

Resultados de aprendizaje

1. Analizar y resolver problemas de forma efectiva, generando propuestas innovadoras y creativas para alcanzar los objetivos.
2. Concebir, diseñar e implementar procesos de recopilación y anotación de datos adecuados al problema concreto a resolver.
3. Conocer los conceptos básicos de sistemas distribuidos de datos y el uso de herramientas de procesamiento masivo de datos.
4. Elegir los métodos de almacenamiento más apropiados que permitan la posterior recuperación y análisis de datos eficiente.
5. Proponer nuevos métodos o soluciones alternativas fundamentadas.
6. Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio.
7. Trabajar cooperativamente para la consecución de objetivos comunes, asumiendo la propia responsabilidad y respetando el rol de los diferentes miembros del equipo.

Contenido

1-Introducción a las aplicaciones masivas de datos

2-Conceptos fundamentales del tratamiento de datos en entornos masivos: fiabilidad, escalabilidad, sostenibilidad. Modelos de datos y lenguajes de consulta.

3-Gestión de grandes volúmenes de datos. Data warehousing. Principios de los sistemas Data Warehousing, Business intelligence, modelización multidimensional, operadores OLAP, esquema en estrella, proceso ETL

4-Introducción a las bases de datos en memoria con Redis

5-Aplicaciones con grandes volúmenes de datos con herramientas Apache. Introducción a las herramientas y ecosistema Apache Spark: Datasets, MLib

Actividades formativas y Metodología

Título	Horas	ECTS	Resultados de aprendizaje
--------	-------	------	---------------------------

Tipo: Dirigidas

Estudio de casos de uso	9	0,36	1, 2, 5, 7
Prácticas	12	0,48	2, 4, 3, 7
Teoría	20	0,8	2, 4, 6, 3
Tipo: Autónomas			
Estudio autónomo	30	1,2	1, 4, 6, 3
Preparación de casos de estudio	20	0,8	1, 2, 5, 4, 3, 7
Preparación de laboratorios	32	1,28	1, 2, 7

En el desarrollo de la asignatura se podrán diferenciar los siguientes tipos de actividades docentes:

Clases teóricas: exposición de la parte teórica de cada tema del programa. La estructura típica de una clase magistral de este tipo será la siguiente: en primer lugar se hará una introducción donde se presentarán brevemente los objetivos de la exposición y los contenidos a tratar. A continuación, se desgranarán los contenidos objeto de estudio, incluyendo exposiciones narrativas, desarrollos formales que proporcionen los fundamentos teóricos, e intercalando ejemplos que ilustren la aplicación de los contenidos expuestos. Finalmente, el/la profesor/a expondrá las conclusiones de los contenidos. Durante todo el curso habrá evaluación continuada de grupos de temas.

Clases en el laboratorio: La parte práctica de los temas teóricos quedará completada con sesiones en el laboratorio, donde los grupos de estudiantes desarrollarán una serie de programas y deberán intentar resolver un problema concreto que recibirá al comenzar el temario. Algunos de estos ejercicios se deberán entregar a la clase en las fechas especificadas. Las prácticas se desarrollarán en grupos de dos estudiantes. Las clases incluyen varias sesiones en el laboratorio, de 2 horas de duración, donde el/la estudiante realizará el desarrollo de los ejercicios.

Problemas: revisión de casos prácticos. Al final de la asignatura se presentará una lista de casos prácticos con retos de datos relacionados con objetivos de análisis de negocio a realizar. Los estudiantes trabajaran en grupo para presentar una serie de conclusiones de su estudio de forma oral en las sesiones finales. Se realizará una evaluación en grupo de los trabajos presentados.

Este planteamiento del trabajo está orientado a promover un aprendizaje activo y desarrollar las competencias de capacidad de organización y planificación, comunicación oral y escrita, trabajo en equipo y razonamiento crítico. La calidad de los ejercicios realizados, de su presentación y de su funcionamiento se valorará especialmente.

Nota: se reservarán 15 minutos de una clase dentro del calendario establecido por el centro o por la titulación para que el alumnado rellene las encuestas de evaluación de la actuación del profesorado y de evaluación de la asignatura o módulo.

Evaluación

Actividades de evaluación continuada

Título	Peso	Horas	ECTS	Resultados de aprendizaje
Control individual 1	30%	2	0,08	1, 6, 3
Control individual 2	30%	2	0,08	1, 6, 3

Estudio de casos prácticos	10%	5	0,2	1, 5, 4, 3, 7
Laboratorios de prácticas	30%	18	0,72	1, 2, 4, 3, 7

El objetivo del proceso de evaluación es verificar que el alumno ha adquirido los conocimientos y habilidades definidos en los objetivos de la asignatura, así como las competencias.

Se evaluarán los siguientes tipos de actividades de manera independiente donde la suma ponderada de ellas dará la nota final.

Teoría (T)

Resolución de las prácticas de laboratorio (PL)

Realización de un informe y una presentación sobre un caso práctico de estudio (PA).

La parte de Teoría (T) se evaluará con dos controles parciales individuales durante todo el curso. La nota final de Teoría saldrá de la suma ponderada de los dos controles ($0.5 * \text{Control 1} + 0.5 * \text{Control 2}$). Habrá una segunda oportunidad para recuperar esta parte el día que tengamos asignado en la semana de exámenes de junio. Solo se podrán recuperar las partes que no hayan sido superadas en los controles parciales de teoría. La nota mínima para aprobar esta parte es $\geq 4,5$. Cada evaluación parcial necesita una nota mínima de 4 para poder hacer media de nota final de teoría. Es necesario aprobar la teoría para poder aprobar la asignatura.

La parte de Resolución de prácticas de laboratorio (PL) se evaluará de manera grupal. Se han planificado inicialmente cuatro entregas. La nota final saldrá de la suma ponderada de las entregas. Para aprobar las PL la nota mínima deberá ser $\geq 4,5$. Sólo hay una única oportunidad (no se puede recuperar esta parte). Es obligatorio entregar los laboratorios de prácticas para aprobar la asignatura.

El estudio de casos prácticos se presentará de forma oral mediante una presentación y escrita mediante un informe. La evaluación se realizará en grupo teniendo en cuenta la valoración cruzada y la corrección por parte del profesor. El valor de esta parte es del 10% de la nota final y dada su naturaleza y objetivo no son recuperables.

La nota final de la asignatura será la suma ponderada de las notas de cada una de las cuatro actividades: 60% de Teoría, 10% Resolución de ejercicios prácticos individuales y 30% de Resolución de prácticas de laboratorio. El resultado deberá ser ≥ 5 .

En caso de no superar la asignatura por no alcanzar la puntuación mínima en alguno de los apartados (Teoría o Prácticas de Laboratorio), aunque al hacer la media ponderada la nota final fuera igual o superior a 5 la nota que se pondrá en el expediente será de 4,5.

En caso de que la media no llegue a 5 la nota que figurará en el expediente será la nota media obtenida numéricamente.

Si el/la alumno/a entrega cualquier actividad, se entiende que se presenta en la asignatura y será evaluado/a. Si no entrega ninguna actividad, entonces se puede considerar No evaluable.

Otorgar una calificación de matrícula de honor es decisión del profesorado responsable de la asignatura. La normativa de la UAB indica que las MH sólo se podrán conceder a estudiantes que hayan obtenido una calificación final igual o superior a 9.00. Se puede otorgar hasta un 5% de MH del total de estudiantes matriculados.

Las fechas de evaluación continua y entrega de trabajos se publicarán en el campus virtual y pueden estar sujetos a cambios de programación por motivos de adaptación a posibles incidencias; siempre se informará en el campus virtual sobre estos cambios ya que se entiende que el CV es el mecanismo habitual de intercambio de información entre el profesorado y los/las estudiantes.

Para cada actividad de evaluación, se indicará un lugar, fecha y hora de revisión en la que el estudiante podrá revisar la actividad con el/la profesor/a. En este contexto, se podrán hacer reclamaciones sobre la nota de la actividad, que serán evaluadas por el profesorado responsable de la asignatura. Si el/la estudiante no se presenta en esta revisión, no se revisará posteriormente esta actividad.

Nota sobre plagios:

Sin perjuicio de otras medidas disciplinarias que se estimen oportunas, y de acuerdo con la normativa académica vigente, las irregularidades cometidas por un estudiante que puedan conducir a una variación de la calificación en una actividad evaluable se calificarán con un cero (0). Las actividades de evaluación calificadas de esta forma y por este procedimiento no serán recuperables. Si es necesario superar cualquiera de estas actividades de evaluación para aprobar la asignatura, esta asignatura quedará suspendida directamente, sin oportunidad de recuperarla en el mismo curso. Estas irregularidades incluyen, entre otros:

la copia total o parcial de una práctica, informe, o cualquier otra actividad de evaluación;

dejar copiar;

presentar un trabajo de grupo no hecho íntegramente por los y las miembros del grupo (aplicado a todos los y las miembros, no sólo a los que no han trabajado);

presentar como propios materiales elaborados por un tercero, aunque sean traducciones o adaptaciones, y en general trabajos con elementos no originales y exclusivos del estudiante;

tener dispositivos de comunicación (como teléfonos móviles, smart watches, bolígrafos con cámara, etc.) accesibles durante las pruebas de evaluación teórico-prácticas individuales (exámenes);

hablar con compañeros o compañeras durante las pruebas de evaluación teórico-prácticas individuales (exámenes);

copiar o intentar copiar de otros alumnos durante las pruebas de evaluación teórico-prácticas (exámenes);

usar o intentar usar escritos relacionados con la materia durante la realización de las pruebas de evaluación teórico-prácticas (exámenes), cuando éstos no hayan sido explícitamente permitidos.

En caso de no superar la asignatura debido a que alguna de las actividades de evaluación no alcanza la nota mínima requerida, la nota numérica del expediente será el valor menor entre 4.5 y la media ponderada de las notas. Con las excepciones de que se otorgará la calificación de "No Evaluable" a los y las estudiantes que no participen en ninguna de las actividades de evaluación, y de que la nota numérica del expediente será el valor menor entre 3.0 y la media ponderada de las notas en caso de que el estudiante haya cometido irregularidades en un acto de evaluación (y por tanto no será posible el aprobado por compensación). En ediciones futuras de esta asignatura, el estudiante que haya cometido irregularidades en un acto de evaluación no se le convalidará ninguna de las actividades de evaluación realizadas.

En resumen: copiar, dejar copiar o plagiar (o el intento de) en cualquiera de las actividades de evaluación equivale a un SUSPENSO, no compensable y sin convalidaciones de partes de la asignatura en cursos posteriores.

Finalmente, los estudiantes que realicen esta asignatura por segunda vez podrán solicitar la convalidación de trabajos prácticos con una nota máxima de 6,0. Si el estudiante obtuvo una calificación inferior a 6,0, se conservará la calificación del año anterior.

Bibliografía

Designing Data intensive applications - Martin Kleppmann, O'Reilly, 2017

The Data warehouse ETL toolkit - Ralph Kimball, Joe Caserta. Wiley, 2004

Spark, the definitive guide, Big data processing made simple. Bill Chambers and Matei Zaharia, O'Reilly, 2018

Learning Spark - Lightning fast data analysis - Holden Karau, Andi Konwinski, Patrick Wendell, Matei Zaharia, O'Reilly, 2015

Beginning Scala - Layka, Vishal. Apress; 2nd ed. 2015.

Redis in Action - Josiah L. Carlson. Manning, 2013.

Software

Se utilizarán los servicios en la nube proporcionados por la Escuela de Ingeniería en la plataforma OpenNebula

Lista de idiomas

Nombre	Grupo	Idioma	Semestre	Turno
(PAUL) Prácticas de aula	1	Inglés	primer cuatrimestre	tarde
(PLAB) Prácticas de laboratorio	1	Inglés	primer cuatrimestre	tarde
(PLAB) Prácticas de laboratorio	2	Inglés	primer cuatrimestre	tarde
(TE) Teoría	1	Inglés	primer cuatrimestre	tarde