

Desarrollo de Aplicaciones de Big Data

Código: 104358
Créditos ECTS: 6

Titulación	Tipo	Curso	Semestre
2503758 Ingeniería de Datos	OB	3	2

La metodología docente y la evaluación propuestas en la guía pueden experimentar alguna modificación en función de las restricciones a la presencialidad que impongan las autoridades sanitarias.

Contacto

Nombre: Antonio Espinosa Morales

Correo electrónico: AntonioMiguel.Espinosa@uab.cat

Uso de idiomas

Lengua vehicular mayoritaria: español (spa)

Algún grupo íntegramente en inglés: No

Algún grupo íntegramente en catalán: No

Algún grupo íntegramente en español: No

Equipo docente externo a la UAB

Carlos Montemuiño

Ramon Grau

Prerequisitos

Aunque no hay prerequisites 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 paradigmas de procesamiento masivo de datos. Modelos de programación para aplicaciones de procesamiento "batch", en memoria y "streaming". Arquitectura de aplicaciones comunes (Lampda, Kappa...). Integridad, accesibilidad, fiabilidad, consistencia y seguridad en el procesamiento de datos a gran escala. Integración de aplicaciones y de infraestructuras: mecanismos de fiabilidad y escalabilidad

Competencias

- Concebir, diseñar e implementar aplicaciones eficientes para el análisis y gestión de datos masivos.
- Concebir, diseñar e implementar sistemas de almacenamiento de datos de forma eficiente y segura.
- Planificar y gestionar el tiempo y los recursos disponibles.
- Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.

Resultados de aprendizaje

1. Aplicar técnicas para automatizar la respuesta de las aplicaciones a situaciones dinámicas (fiabilidad, escalabilidad, emergencias, ...).
2. Desarrollar aplicaciones que procesen datos a gran escala usando paradigmas batch y streaming.
3. Planificar y gestionar el tiempo y los recursos disponibles.
4. Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.

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. Formatos de codificación de la información
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. Aplicaciones con grandes volúmenes de datos con herramientas Apache. Introducción a las herramientas y ecosistema Apache Hadoop. Gestión de los datos de entrada y procesos ETL, tipos de datos, introducción a Apache Spark para el procesamiento de grandes volúmenes de datos.

Metodología

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 prácticas: Todos los temas irán acompañados de un lista de problemas que el alumno debe intentar resolver. En este sentido, y a medida que el alumno vaya progresando y profundizando en sus conocimientos, estos problemas serán poco a poco más complejos. Los seminarios serán el foro natural en el que se podrá discutir en común el desarrollo del trabajo práctico, aportando los conocimientos que le faltan al estudiante para llevarlo adelante. La misión de las clases prácticas es hacer de puente entre las clases teóricas y las clases de laboratorio, que promoverá la capacidad de análisis y síntesis, el razonamiento crítico, y que entrenará al estudiante en la resolución de problemas. Aquellos que el/la profesor/a considere de mayor interés o en los que los alumnos/as encuentren mayor dificultad serán corregidos en la pizarra. Antes del comienzo de cada tema de problemas el/la profesor/a podrá proponer una lista de ejercicios que los alumnos/as deberán resolver y entregar en el campus virtual para ser evaluados.

Clases en el laboratorio: La parte práctica de los temas teóricos quedará completada con sesiones en el laboratorio, donde el/la alumno/a desarrollará una serie de programas y deberá 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 o tres alumnos. Las clases incluyen 6 sesiones en el laboratorio, de 2 horas de duración, donde el/la alumno/a realizará el desarrollo de los ejercicios. El resultado de estas sesiones será un informe de prácticas que se deberá entregar en el campus virtual para ser evaluado.

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.

Actividades

Título	Horas	ECTS	Resultados de aprendizaje
Tipo: Dirigidas			
Problemas	9	0,36	1, 3, 4
Prácticas	12	0,48	2, 3, 4
Teoría	20	0,8	3, 4
Tipo: Autónomas			
Estudio autónomo	30	1,2	3, 4
Preparación de ejercicios	20	0,8	1, 3, 4
Preparación de laboratorios	32	1,28	2, 3, 4

Evaluación

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 ejercicios prácticos individuales (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 ($0.5 \cdot \text{Control 1} + 0.5 \cdot \text{Control 2}$). Habrá una segunda oportunidad para recuperar esta parte el día que tengamos asignado en la semana de exámenes de junio. 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.

La parte de Resolución de prácticas de laboratorio (PL) se evaluará de manera grupal. Tiene dos 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.

Los ejercicios prácticos individuales (PA) consistirán en trabajar problemas de programación muy concretos y en algunos casos relacionados con los que se encontrarán en las prácticas de laboratorio. Se quiere que cada

estudiante estudie la resolución de un conjunto de problemas concretos de forma aislada al caso más general de la práctica. El valor de estos ejercicios 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.

Actividades de evaluación

Título	Peso	Horas	ECTS	Resultados de aprendizaje
Control individual parcial 1	30%	2	0,08	1, 3, 4
Control individual parcial 2	30%	2	0,08	1, 3, 4
Entrega prácticas	30%	18	0,72	2, 3, 4
Resolución ejercicios prácticos individuales	10%	5	0,2	1, 3, 4

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.

Software

Talend Open Studio for Data Integration

MongoDB

Redis

Apache Spark