

**FICHA IDENTIFICATIVA****Datos de la Asignatura**

Código	34695
Nombre	Sistemas de Información de nueva generación
Ciclo	Grado
Créditos ECTS	6.0
Curso académico	2023 - 2024

Titulación(es)

Titulación	Centro	Curso	Periodo
1400 - Grado de Ingeniería Informática	Escuela Técnica Superior de Ingeniería	4	Segundo cuatrimestre

Materias

Titulación	Materia	Caracter
1400 - Grado de Ingeniería Informática	16 - Materia Optativa	Optativa

Coordinación

Nombre	Departamento
FUERTES SEDER, ARIADNA	240 - Informática
VES CUENCA, ESTHER DE	240 - Informática

RESUMEN

El enorme volumen de datos que se almacenan hoy en día está creciendo de forma exponencial. En el año 2000 la información almacenada era del orden de los 800.000 petabytes (PB). Se estima que esta cifra alcanzará los 35 zetabytes (ZB) en 2020. Sólo Twitter genera más de 7 terabytes (TB) de datos cada día, Facebook 10 TB y algunas empresas generan terabytes de datos cada hora de cada día del año. Ya no es extraño encontrar empresas que cuentan con capacidades de almacenamiento de datos del orden del petabyte. Los motores de este crecimiento son principalmente el desarrollo de la Web, el número creciente de proveedores de redes sociales, el desarrollo de los servicios en la nube y en general el modelo de “software como servicio”.

Paralelamente, la capacidad para tratar y explorar toda esta sobrecarga de información y aplicar las técnicas de extracción de significado útil son críticas en cualquier organización, empresa e institución actual. No es extraño, por tanto, que la demanda de sistemas de alta escalabilidad haya crecido también de forma exponencial para adaptarse al volumen de datos y a las necesidades de procesamiento.



Aunque la bases de datos relacionales han demostrado su capacidad para adaptarse con éxito a la creciente necesidad de almacenamiento, ahora más que nunca son necesarios nuevos paradigmas de persistencia sólidos, fiables y escalables capaces de atender las demandas de los nuevos sistemas de información. Son dos las estrategias que se están adoptando para resolver este problema:

1. Por un lado, los sistemas relacionales tradicionales se pueden adaptar para responder a las necesidades de escalabilidad mediante la paralelización (mayor velocidad), la distribución (mayor capacidad de almacenamiento y proceso) y la federación y replicación (escalabilidad horizontal mediante la simple adición de nodos).
2. Por otro lado, la introducción de nuevos paradigmas de bases de datos, tales como NoSQL o Big Data, que intentar atacar el problema proponiendo nuevas estructuras de almacenamiento más versátil, aunque sea a costa de la pérdida de ciertas funcionalidades tradicionalmente consideradas como buenas.

En esta asignatura se estudian y analizan las ventajas e inconvenientes de ambas estrategias y se presentan las nuevas tecnologías de persistencia que el profesional de las Tecnologías de la Información y la Comunicación (TIC) debe necesariamente conocer para integrarse en el mercado laboral actual.

CONOCIMIENTOS PREVIOS

Relación con otras asignaturas de la misma titulación

No se han especificado restricciones de matrícula con otras asignaturas del plan de estudios.

Otros tipos de requisitos

Para una adecuada comprensión de la asignatura es necesario haber cursado las asignaturas Bases de Datos de 2º curso y Sistemas de Gestión de Bases de Datos de 3º curso.

COMPETENCIAS

1400 - Grado de Ingeniería Informática

- SI1 - Capacidad de integrar soluciones de Tecnologías de la Información y las Comunicaciones y procesos empresariales para satisfacer las necesidades de información de las organizaciones, permitiéndoles alcanzar sus objetivos de forma efectiva y eficiente, dándoles así ventajas competitivas.
- SI3 - Capacidad para participar activamente en la especificación, diseño, implementación y mantenimiento de los sistemas de información y comunicación.



RESULTADOS DE APRENDIZAJE

- • Comprender la motivación de los sistemas de gestión de bases de datos paralelos, distribuidos y federados.
- • Disponer de los fundamentos básicos para analizar las diferentes arquitecturas, las implicaciones en el rendimiento, la velocidad, la escalabilidad y la seguridad de las diferentes estrategias de particionado de los datos y evaluar y optimizar las consultas sobre datos distribuidos.
- • Comprender qué es data warehousing y por qué es importante en el soporte para la toma de decisiones.
- • Comprender el modelo de datos multidimensional y el tipo de análisis de datos que facilita.
- • Comprender qué es data mining y disponer de los fundamentos básicos de las técnicas utilizadas en data mining.
- • Conocer el algoritmo MapReduce y su utilización en el análisis rápido de cantidades inmensas de datos.
- • Conocer los diferentes sistemas de almacenamiento masivo de información conocidos genéricamente como sistemas NoSQL y disponer de los fundamentos básicos para definir, diseñar e implementar sistemas de información utilizando estos sistemas.

DESCRIPCIÓN DE CONTENIDOS

1. Sistemas avanzados de gestión de bases de datos

Bases de datos distribuidas.
Bases de datos paralelas.
Federación de bases de datos.
Replicación.

2. Sistemas de apoyo para la toma de decisiones

Almacenes de datos (Data warehouses)
Minería de Datos (Data mining)

3. Big data y NoSQL

Terminología
Bases de datos NoSQL
MapReduce
Almacenamiento distribuido

**4. Recuperación de información y datos XML**

Recuperación de información.

Modelo de datos para XML.

XQuery.

Evaluación eficiente de consultas XML.

VOLUMEN DE TRABAJO

ACTIVIDAD	Horas	% Presencial
Clases de teoría	30,00	100
Prácticas en laboratorio	20,00	100
Prácticas en aula	10,00	100
Elaboración de trabajos en grupo	5,00	0
Elaboración de trabajos individuales	20,00	0
Estudio y trabajo autónomo	10,00	0
Lecturas de material complementario	5,00	0
Preparación de actividades de evaluación	15,00	0
Preparación de clases de teoría	10,00	0
Preparación de clases prácticas y de problemas	25,00	0
TOTAL	150,00	

METODOLOGÍA DOCENTE

La docencia consistirá en una combinación de lecciones teóricas, sesiones de problemas y actividades de carácter práctico a realizar por parte del estudiante. Esta docencia quedará complementada con el trabajo personal del alumnado, centrado en el estudio, en la resolución de problemas, y en la preparación de trabajos para entregar. Además, se realizarán sesiones de laboratorio basadas en trabajo con el ordenador.

- Las actividades teóricas consistirán en la realización de clases de una hora y media de duración en las que se desarrollarán los temas proporcionando una visión global e integradora, analizando con mayor detalle los aspectos clave y de mayor complejidad, fomentando, en todo momento, la participación del alumnado.
- Las actividades prácticas consistirán en la realización de sesiones de problemas y cuestiones en el aula de hora y media de duración, así como la realización de seminarios, en los que se abordarán dos temas sobre aplicaciones y aspectos menos formales de la asignatura. Complementan las lecciones expositivas con el objetivo de aplicar los conceptos básicos y ampliarlos con el conocimiento y la experiencia que vayan adquiriendo durante la realización de los trabajos propuestos. Como ejemplos de estas actividades podemos mencionar las siguientes:
 - o Clases de problemas y cuestiones en aula
 - o Seminarios aplicados en pizarra y aula de ordenadores
 - o Sesiones de discusión y resolución de problemas y ejercicios previamente trabajados por



- los estudiantes
 - o Tutorías individualizadas
- Las sesiones de laboratorio consistirán en la resolución de problemas relacionados con los contenidos teóricos por medio de programas informáticos en sesiones de dos horas y media de duración.
- El trabajo personal del alumnado intenta fomentar el trabajo autónomo y consistirá, fundamentalmente, en tres aspectos:
 - o La preparación de las clases con antelación y la lectura de textos recomendados y preparación de exámenes (estudio).
 - o La resolución de problemas propuestos por el profesor
 - o La elaboración de trabajos que se entregarán al profesor.
 - o La preparación de las sesiones de laboratorio con anterioridad a la asistencia al aula de prácticas informáticas.
- El Trabajo en pequeños grupos de estudiantes (2-4) para la realización de actividades, cuestiones, problemas fuera del aula. Esta tarea complementa el trabajo individual y fomenta la capacidad de integración en grupos de trabajo.

Se utilizará la plataforma de e-learning (Aula Virtual) de la Universitat de València como soporte de comunicación con los estudiantes. A través de ella se tendrá acceso al material didáctico utilizado en clase, así como los problemas y ejercicios a resolver.

EVALUACIÓN

El trabajo que se espera que los alumnos realicen lo podemos clasificar en 2 tipos:

1. Trabajo autónomo de auto-aprendizaje.

2. Trabajo supervisado.

1. El trabajo autónomo de auto-aprendizaje consiste en las actividades que el alumno realiza fuera del horario de docencia, de manera independiente o dirigida por el profesor, y que está orientado a adquirir conocimientos, capacidades y destrezas de forma autodidacta. Principalmente serán actividades que el profesor indique que se deben realizar durante el curso (lecturas recomendadas, realización de problemas, investigación de ciertos temas, etc.) pero que no serán evaluadas directamente (calificadas) por el profesor, aunque tendrán incidencia en otras actividades que sí que serán evaluadas y calificadas por el profesor, como son el caso de los debates y presentaciones realizados durante las horas de docencia o la realización de exámenes.

2. El trabajo supervisado consiste en todas las actividades que realizará el alumno a petición del profesor y que éste monitorizará para evaluar el grado de superación de la asignatura. Estos trabajos serán de 3 tipos:

a) Problemas o actividades Individuales

b) En grupo

c) Prácticas de aula de informática, de asistencia obligatoria.

Las características de estos trabajos son:



- Deben ser evaluadas por el profesor.
- Deben tener fecha de entrega, y/o realizarse de manera presencial en el momento de ser planteadas, lo que obligaría a la asistencia en dichos casos.
- Son de realización obligatoria por parte del alumno para la aplicación de la evaluación continua.

En primera convocatoria, el alumno realizará a lo largo del curso varias pruebas objetivas individuales durante las horas lectivas y que consistirán en tanto cuestiones teórico-prácticas como problemas o debates (N_Pruebas) que no serán recuperables.

Además, habrá que entregar distintas actividades correspondientes a las actividades desarrolladas en el laboratorio (N_Actividades).

Existirá finalmente un examen (N_Examen).

La calificación final, en primera convocatoria, se obtendrá mediante la siguiente ponderación de las distintas actividades y pruebas:

$$\text{Nota Final} = 20\% \text{ N_Examen} + 40\% \text{ N_Pruebas} + 40\% \text{ N_Actividades}$$

En segunda convocatoria la calificación final se obtendrá en base a la fórmula:

$$\text{Nota Final} = 60\% \text{ N_Examen} + 40\% \text{ N_Actividades}$$

Será necesario obtener, al menos, 5 sobre 10 en cada una de las partes para poder mediar la nota.

En cualquier caso, la evaluación de la asignatura se hará de acuerdo con el Reglamento de evaluación y calificación de la Universitat de València para los títulos de grado y master aprobado por Consejo de Gobierno de 30 de mayo de 2017 (ACGUV 108/2017).

Esta evaluación parte de la premisa de que la docencia en la Universitat de València es, por definición, una docencia presencial. En este sentido, el alumno debe tener presente que la asistencia, tanto a las clases teóricas como a aquellas de carácter práctico, es fundamental para un adecuado seguimiento de los contenidos de la asignatura. El alumno debe tener presente igualmente la posibilidad de una matrícula a tiempo parcial, salvo en el caso de los alumnos de 1º, cuando no le sea posible asistir a la totalidad de las asignaturas que componen un curso completo (60 créditos). Con todo, se establecerá la posibilidad, en los casos que estén adecuadamente justificados y para aquellos alumnos que lo soliciten, la posibilidad de ser evaluado sin necesidad de asistir a la totalidad o a parte de las clases. En estos casos el alumno debe proceder del siguiente modo:



· Se debe comunicar al principio del curso al profesor/es responsable/s de la asignatura la incidencia por la que le es imposible asistir a clase, que debe estar adecuadamente justificada de forma documental.

· El profesor responsable, a la vista de esta información decidirá la posibilidad de evaluación sin asistencia total o parcial a las clases de la asignatura.

Los alumnos que se encuentren en esta situación, deberán presentar, para ser evaluados, la totalidad de trabajos requeridos por el profesor (no necesariamente idénticos a los requeridos durante el curso) así como también podrán ser llamados a defenderlos oralmente ante el propio profesor, y realizarán una prueba de conocimientos adquiridos. El peso de los trabajos en la calificación final será de un 50% y el de la prueba de conocimientos el 50% restante.

REFERENCIAS

Básicas

- Ramez A. Elmasri, Shmkant B. Navathe Fundamentos de Sistemas de Bases de Datos. Addison Wesley, 2002.
- R. Ramakrishnan, J. Gehrke Database Management Systems. McGraw-Hill, 2000.
- S. Tiwari. Professional NoSQL. Wrox, 2011.
- T. White. Hadoop: the definitive guide. Wrox, 2009

Complementarias

- Referencia c1: Priscilla Walmsley. XQuery. O'Reilly Media, Inc., 2007
- H. Inmon. Building the Data Warehouse, John Wiley & Sons, 2005
- D. Miner. MapReduce design patterns. O'Reilly 2012
- Christof Strauch. NoSQL databases. <http://www.christof-strauch.de/nosql dbs.pdf>
- C. Adamson. Mastering data warehouse aggregates. Willey, 2006