

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

Código	36424
Nombre	Gestión de datos
Ciclo	Grado
Créditos ECTS	6.0
Curso académico	2021 - 2022

Titulación(es)

Titulación	Centro	Curso	Periodo
1406 - Grado en Ciencia de Datos	Escuela Técnica Superior de Ingeniería	2	Primer cuatrimestre

Materias

Titulación	Materia	Caracter
1406 - Grado en Ciencia de Datos	8 - Gestión de la información	Obligatoria

Coordinación

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

RESUMEN

La asignatura “Gestión de datos” forma parte de la materia “Gestión de la información” y es una asignatura de 6 créditos ECTS que se imparte en el primer cuatrimestre de segundo curso del Grado en Ciencia de Datos.

El volumen de datos que se genera, almacena y procesa hoy en día crece de forma imparable. Aunque las bases de datos relacionales han demostrado su capacidad para adaptarse con éxito a las crecientes necesidades, las arquitecturas y modelos tradicionales de bases de datos en ocasiones pueden resultar insuficientes o inadecuadas para tratar el volumen y tipo de datos que se quieren gestionar.

Esta asignatura trata tanto nuevas arquitecturas de bases de datos paralelas y distribuidas, como nuevos modelos de bases de datos generalmente conocidos como NoSQL, así como herramientas y técnicas de almacenes de datos (Data Warehousing) y modelos multidimensionales de datos, junto a una visión general de los procesos de integración y migración de datos.



La asignatura proporcionará las capacidades de análisis de las ventajas e inconvenientes de las diferentes tecnologías para ser capaz de seleccionar las más adecuadas en cada contexto.

Las clases de teoría se impartirán en castellano y las clases prácticas y de laboratorio según consta en la ficha de la asignatura disponible en la web del grado.

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 un adecuado aprendizaje en la asignatura es necesario haber cursado con aprovechamiento la asignatura Bases de Datos de primer curso del Grado en Ciencia de Datos.

COMPETENCIAS

1406 - Grado en Ciencia de Datos

- (CG03) Capacidad para la realización de modelos, cálculos, informes, planificación de tareas y otros trabajos análogos en el ámbito específico de la Ciencia de Datos.
- (CG06) Capacidad de acceso y gestión de la información en diferentes formatos para su posterior análisis con el fin de obtener conocimiento a partir de datos.
- (CT02) Ser capaces de completar su formación técnica, científica, social y humana en general, y de organizar su propio autoaprendizaje con un alto grado de autonomía.
- (CT05) Capacidad para evaluar las ventajas e inconvenientes de diferentes alternativas metodológicas y/o tecnológicas en distintos ámbitos de aplicación.
- (CE02) Conocer y aplicar de forma metodológica las técnicas de programación y la algoritmia necesarias para el procesamiento eficiente de información y la resolución informática de problemas que utilizan grandes volúmenes de datos.
- (CE04) Conocer y utilizar los distintos modelos de almacenamiento de datos y los sistemas de gestión de las bases de datos utilizando lenguajes de programación de definición, consulta y manipulación de los mismos.
- (CE08) Capacidad para comprender, seleccionar y utilizar la infraestructura y técnicas adecuadas para el tratamiento de datos masivos, atendiendo a criterios de eficiencia, escalabilidad, seguridad, tolerancia a fallos y adecuación al entorno de producción.
- (CE11) Capacidad para diseñar e implementar la toma de datos, su integración, transformación, selección, comprobación de su calidad y veracidad a partir de distintas fuentes, teniendo en cuenta su carácter, heterogeneidad y variabilidad.



- (CB4) Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.

RESULTADOS DE APRENDIZAJE

Comprender la motivación de los sistemas de gestión de bases de datos paralelos y distribuidos. (CT02)

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. (CB4, CG03, CE08)

Comprender qué es un almacén de datos y la razón de su importancia como soporte a la toma de decisiones. (CT05, CE02)

Conocer el modelo de datos multidimensional y el tipo de análisis de datos que facilita. (CE02)

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. (CT05, CE02, CE04, CE08)

Saber cómo realizar un proceso de integración de datos de distintas fuentes y migración. (CG06, CE11)

DESCRIPCIÓN DE CONTENIDOS

1. Bases de datos paralelas y distribuidas

Arquitecturas avanzadas de sistemas de bases de datos. Bases de datos paralelas. Procesamiento de consultas en paralelo. Bases de datos distribuidas. Almacenamiento, procesamiento y recuperación en bases de datos distribuidas.

2. Evaluación y optimización de consultas en diversas arquitecturas

Evaluación de las operaciones relacionales: selección, proyección, reunión, operaciones de consulta y agregación. Optimización de consultas: estimación del coste de un plan, equivalencias, algoritmos para la optimización de consultas, determinación del plan físico. Evaluación de consultas en bases de datos paralelas y distribuidas.



3. Almacenes de datos y modelo multidimensional de datos

Almacenes de datos: motivación, estructura y características. Modelo multidimensional de datos: modelos OLAP, operaciones, SQL:1999.

4. Bases de datos NoSQL

Modelos de datos adaptados a nuevos requerimientos: datos a gran escala, información semiestructurada y no estructurada, arquitecturas distribuidas. Bases de datos orientadas a columnas. Bases de datos orientadas a documentos. Bases de datos orientadas a grafos.

5. Integración de datos y procesos de migración de datos

Integración de datos de diferentes fuentes. El proceso ETL: extracción, transformación y carga. Aspectos a considerar en la migración de datos.

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 individuales	15,00	0
Estudio y trabajo autónomo	10,00	0
Lecturas de material complementario	15,00	0
Preparación de actividades de evaluación	10,00	0
Preparación de clases de teoría	10,00	0
Preparación de clases prácticas y de problemas	30,00	0
TOTAL	150,00	

METODOLOGÍA DOCENTE

MD1 - Actividades teóricas. Desarrollo expositivo de la materia con la participación del estudiante en la resolución de cuestiones puntuales. Realización de pruebas individuales de evaluación. (CG03, CT02, CT05, CE04, CE08)

MD2 - Actividades prácticas. Aprendizaje mediante resolución de problemas, ejercicios y casos de estudio a través de los cuales se adquieren competencias sobre los diferentes aspectos de la materia. (CB4, CE08, CE11)



MD4 - Trabajos en laboratorio y/o aula ordenador. Aprendizaje mediante la realización de actividades desarrolladas de forma individual o en grupos reducidos y llevadas a cabo en laboratorios y/o aulas de ordenador. (CG06, CE02, CE04, CE08)

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

EVALUACIÓN

SE1 - Prueba objetiva, consistente en uno o varios exámenes que constan tanto de cuestiones teórico-prácticas como de problemas. La nota obtenida en esta(s) prueba(s) representará el 50% de la nota final de la primera convocatoria, y además será necesaria una nota mínima de 5 puntos sobre 10 en este bloque para superar la asignatura. (CG03, CT02, CT05, CE04, CE08)

SE2 - Evaluación de las actividades prácticas a partir de la elaboración de trabajos/memorias y/o exposiciones orales. Estas actividades prácticas se realizarán en el laboratorio de informática y son de obligada asistencia excepto causa adecuadamente justificada. La nota obtenida en este bloque representará el 30% de la nota final. (CB4, CE08, CE11)

SE3 - Evaluación continua de cada alumno, basada en la participación y grado de implicación del alumno en el proceso de enseñanza-aprendizaje, teniendo en cuenta la asistencia regular a las actividades presenciales previstas y la resolución de cuestiones y problemas propuestos periódicamente, así como la presentación y exposición de trabajos. La nota obtenida en este bloque representará el 20% de la nota final. (CG06, CE02, CE04, CE08)

Las actividades correspondientes a los bloques SE2 y SE3 no son recuperables.

En segunda convocatoria se realizará un examen que representará el 70% de la nota final, siendo necesario un mínimo de 5 puntos sobre 10 en dicho examen, y la nota obtenida durante el periodo lectivo en el bloque SE2 supondrá el 30% restante.

En cualquier caso, el sistema de evaluación se regirá por lo establecido en el Reglamento de Evaluación y Calificación de la Universidad de Valencia para Grados y Masters

(<https://webges.uv.es/uvTaeWeb/MuestraInformacionEdictoPublicoFrontAction.do?accion=inicio&idEdictoSeleccionado=5639>)



REFERENCIAS

Básicas

- Ramakrishnan, R.; Gehrke, J. Sistemas de gestión de bases de datos / Database Management Systems. McGraw-Hill
- Elmasri, R.; Navathe S. B. Fundamentos de bases de datos / Fundamentals of Database System. Pearson
- Silberschatz, A.; Korth, H. F.; Sudarshan, S. Fundamentos de bases de datos / Database System Concepts. McGraw-Hill

Complementarias

- Inmon, W. H.; Linstedt, Daniel. Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault. Morgan Kaufmann Publishers Inc
- Adamson, C.; Kimball, R. Mastering data warehouse aggregates. Willey
- Strauch, Christof. NoSQL databases. Stuttgart Media University.
<http://www.christof-strauch.de/nosql dbs.pdf>
- Robinson, Ian; Webber, Jim; Eifrem, Emil. Graph Databases. O'Reilly Media

ADENDA COVID-19

Esta adenda solo se activará si la situación sanitaria lo requiere y previo acuerdo del Consejo de Gobierno

La metodología docente de la asignatura seguirá el Modelo Docente aprobado por la Comisión Académica de Título de Ciencia de Datos (<https://go.uv.es/cienciadatos/ModelDocentGCD>).

En caso de que se produzca un cierre de las instalaciones por causas sanitarias que afecte total o parcialmente a las clases de la asignatura, estas serán sustituidas por sesiones no presenciales siguiendo los horarios establecidos. Si el cierre afectara a alguna prueba de evaluación presencial de la asignatura, esta será sustituida por una prueba de naturaleza similar que se realizará en modalidad virtual a través de las herramientas informáticas soportadas por la Universitat de València. Los porcentajes de cada prueba de evaluación permanecerán invariables, según lo establecido por esta guía.