

**FITXA IDENTIFICATIVA****Dades de l'Assignatura**

Codi	36424
Nom	Gestió de dades
Cicle	Grau
Crèdits ECTS	6.0
Curs acadèmic	2020 - 2021

Titulació/titulacions

Titulació	Centre	Curs	Període
1406 - Grau en Ciència de Dades	Escola Tècnica Superior d'Enginyeria	2	Primer quadrimestre

Matèries

Titulació	Matèria	Caràcter
1406 - Grau en Ciència de Dades	8 - Gestió de la informació	Obligatòria

Coordinació

Nom	Departament
CERVERON LLEO, VICENTE	240 - Informàtica
FUERTES SEDER, ARIADNA	240 - Informàtica
VES CUENCA, ESTHER DE	240 - Informàtica

RESUM

L'assignatura “Gestió de dades” forma part de la matèria “Gestió de la informació” i és una assignatura de 6 crèdits ECTS que s'imparteix en el primer quadrimestre de segon curs del Grau en Ciència de Dades.

El volum de dades que es creen, emmagatzemen i processen actualment creix de forma imparable. Encara que les bases de dades relacionals han demostrat la seua capacitat per a adaptar-se a les necessitats creixents, les arquitectures i els models tradicionals de bases de dades de vegades poden resultar insuficients o inadequades per a tractar el volum i els tipus de dades que es volen gestionar.

Aquesta assignatura tracta noves arquitectures de bases de dades paral·leles i distribuïdes, també nous models de bases de dades coneguts generalment com NoSQL, així com eines i tècniques de magatzems de dades (Data Warehousing) i models multidimensionals de dades, junt amb una visió general dels processos d'integració i migració de dades.



L'assignatura proporcionarà les capacitats d'anàlisis dels avantatges i els inconvenients de les diferents tecnologies per a ser capaç de seleccionar les més adequades en cada context.

Les classes de teoria s'impartiran en castellà i les classes pràctiques i de laboratori segons el que conste en la fitxa de l'assignatura disponible en la web del grau.

CONEIXEMENTS PREVIS

Relació amb altres assignatures de la mateixa titulació

No heu especificat les restriccions de matrícula amb altres assignatures del pla d'estudis.

Altres tipus de requisits

Per a un bon aprenentatge en l'assignatura és necessari haver cursat amb aprofitament l'assignatura Bases de dades de primer curs del Grau en Ciència de Dades.

COMPETÈNCIES

1406 - Grau en Ciència de Dades

- Que els estudiants puguin transmetre informació, idees, problemes i solucions a un públic tant especialitzat com no especialitzat.
- (CG03) Capacitat per a la realització de models, càlculs, informes, planificació de tasques i altres treballs anàlegs en l'àmbit específic de la Ciència de Dades.
- (CG06) Capacitat d'accés i gestió de la informació en diferents formats per a la seva posterior anàlisi amb la finalitat d'obtenir coneixement a partir de dades.
- (CT02) Ser capaços de completar la seva formació tècnica, científica, social i humana en general, i d'organitzar el seu propi autoaprenentatge amb un alt grau d'autonomia.
- (CT05) Capacitat per avaluar els avantatges i inconvenients de diferents alternatives metodològiques i/o tecnològiques en diferents àmbits d'aplicació.
- (CE02) Conèixer i aplicar de forma metodològica les tècniques de programació i l'algorísmia necessàries per al processament eficient d'informació i la resolució informàtica de problemes que utilitzen grans volums de dades.
- (CE04) Conèixer i utilitzar els diferents models d'emmagatzematge de dades i els sistemes de gestió de les bases de dades utilitzant llenguatges de programació de definició, consulta i manipulació dels mateixos.
- (CE08) Capacitat per comprendre, seleccionar i utilitzar la infraestructura i tècniques adequades per al tractament de dades massives, atenent a criteris d'eficiència, escalabilitat, seguretat, tolerància a fallades i adequació a l'entorn de producció.



- (CE11) Capacitat per dissenyar i implementar la presa de dades, la seva integració, transformació, selecció, comprovació de la seva qualitat i veracitat a partir de diferents fonts, tenint en compte el seu caràcter, heterogeneïtat i variabilitat.

RESULTATS DE L'APRENTATGE

Comprendre la motivació dels sistemes de gestió de bases de dades paral·leles i distribuïdes. (CT02)

Disposar dels fonaments bàsics per a analitzar les diferents arquitectures, les implicacions quant a rendiment, velocitat, escalabilitat i seguretat de les diferents estratègies de partició de les dades i avaluar i optimitzar les consultes sobre dades distribuïdes. (CB4, CG03, CE08)

Comprendre què és un magatzem de dades i la raó de la seua importància com a suport per a la presa de decisions. (CT05, CE02)

Conèixer el model de dades multidimensional i el tipus d'anàlisi de dades que facilita. (CE02)

Conèixer els diferents sistemes d'emmagatzematge massiu d'informació coneguts genèricament com sistemes NoSQL i disposar dels fonaments bàsics per a definir, dissenyar i posar en marxa sistemes d'informació utilitzant aquests sistemes. (CT05, CE02, CE04, CE08)

Saber com dur un procés d'integració de dades de diferents fonts i migració. (CG06, CE11)

DESCRIPCIÓ DE CONTINGUTS

1. Bases de dades paral·leles i distribuïdes

Arquitectures avançades de sistemes de bases de dades. Bases de dades paral·leles. Processament de consultes en paral·lel. Bases de dades distribuïdes. Emmagatzematge, processament i recuperació en bases de dades distribuïdes.

2. Avaluació i optimització de consultes en diferents arquitectures

Avaluació de les operacions relacionals: selecció, projecció, join, operacions de consulta i agregació. Optimització de consultes: estimació del cost d'un pla, equivalències, algorismes per a l'optimització de consultes, determinació del pla físic. Avaluació de consultes en bases de dades paral·leles i distribuïdes.



3. Magatzems de dades i model multidimensional de dades

Magatzems de dades: motivació, estructura i característiques. Model multidimensional de dades: models OLAP, operacions, SQL:1999.

4. Bases de dades NoSQL

Models de dades adaptades a nous requeriments: dades a gran escala, informació semiestructurada i no estructurada, arquitectures distribuïdes. Bases de dades orientades a columnes. Bases de dades orientades a documents. Bases de dades orientades a grafs.

5. Integració de dades i processos de migració de dades

Integració de dades de diferents fonts. El procés ETL: extracció, transformació i càrrega. Aspectes que cal considerar en la migració de dades.

VOLUM DE TREBALL

ACTIVITAT	Hores	% Presencial
Classes de teoria	30,00	100
Pràctiques en laboratori	20,00	100
Pràctiques en aula	10,00	100
Elaboració de treballs individuals	15,00	0
Estudi i treball autònom	10,00	0
Lectures de material complementari	15,00	0
Preparació d'activitats d'avaluació	10,00	0
Preparació de classes de teoria	10,00	0
Preparació de classes pràctiques i de problemes	30,00	0
TOTAL	150,00	

METODOLOGIA DOCENT

MD1 - Activitats teòriques. Desenvolupament expositiu de la matèria amb la participació de l'estudiant en la resolució de qüestions puntuals. Realització de proves individuals d'avaluació. (CG03, CT02, CT05, CE04, CE08)

MD2 - Activitats pràctiques. Aprenentatge mitjançant resolució de problemes, exercicis i casos d'estudi a través dels quals s'adquireixen competències sobre els diferents aspectes de la matèria. (CB4, CE08, CE11)



MD4 - Treballs en laboratori i/o aula ordinador. Aprenentatge mitjançant la realització d'activitats individuals o en grups reduïts als laboratoris i/o a les aules d'ordinador. (CG06, CE02, CE04, CE08)

S'utilitzarà la plataforma d'aprenentatge en línia (Aula Virtual) de la Universitat de València com a suport de comunicació amb l'alumnat, en la qual estan disponibles al material didàctic utilitzat en classe i els problemes i exercicis que cal resoldre.

AVALUACIÓ

SE1 - Prova objectiva, consistent en un o diferents exàmens que consten tant de qüestions teòricopràctiques com de problemes. La nota obtinguda en aquesta(es) prova(es) representarà el 50% de la nota final de la primera convocatòria. En aquest bloc serà necessària una nota mínima de 5 punts sobre 10 per a aprovar l'assignatura. (CG03, CT02, CT05, CE04, CE08)

SE2 - Avaluació de les activitats pràctiques a partir de l'elaboració de treballs/memòries i/o exposicions orals. Aquestes activitats pràctiques es realitzaran al laboratori d'informàtica. L'assistència és obligatòria, excepte si hi ha causa adequadament justificada. La nota obtinguda en aquest bloc representarà el 30% de la nota final. (CB4, CE08, CE11)

SE3 - Avaluació contínua de cada alumne, basada en la participació i el grau d'implicació de l'alumne en el procés d'ensenyament-aprenentatge, tenint en compte l'assistència regular a les activitats presencials previstes, la resolució de qüestions i problemes proposats periòdicament, així com la presentació i l'exposició de treballs. La nota obtinguda en aquest bloc representarà el 20% de la nota final. (CG06, CE02, CE04, CE08)

Les activitats corresponents als blocs SE2 i SE3 no són recuperables.

En la segona convocatòria es farà un examen que representarà el 70% de la nota final, i en què serà necessari un mínim de 5 punts sobre 10. La nota obtinguda durant el període lectiu en el bloc SE2 suposarà el 30% restant.

En qualsevol cas, el sistema d'avaluació es regirà pel Reglament d'Avaluació i Qualificació de la Universitat de València per a Graus i Màsters



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

REFERÈNCIES

Bàsiques

- 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

Complementàries

- 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

ADDENDA COVID-19

Aquesta addenda només s'activarà si la situació sanitària ho requereix i previ acord del Consell de Govern

La metodologia docent de l'assignatura seguirà el Model Docent aprovat per la Comissió Acadèmica de Títol de Ciència de Dades (<https://go.uv.es/cienciadatos/ModelDocentGCD>). En cas que es produísca un tancament de les instal·lacions per causes sanitàries que afecte total o parcialment a les classes de l'assignatura, aquestes seran substituïdes per sessions no presencials seguint els horaris establerts. Si el tancament afectara a alguna prova d'avaluació presencial de l'assignatura, aquesta serà substituïda per una prova de naturalesa similar que es realitzarà en modalitat virtual a través de les eines informàtiques suportades per la Universitat de València. Els percentatges de cada prova d'avaluació romandran invariables, segons allò establert per aquesta guia.