

Guia docent

270087 - CBDE - Conceptes per a Bases de Dades Especialitzades

Última modificació: 11/07/2024

Unitat responsable: Facultat d'Informàtica de Barcelona
Unitat que imparteix: 747 - ESSI - Departament d'Enginyeria de Serveis i Sistemes d'Informació.
Titulació: GRAU EN ENGINYERIA INFORMÀTICA (Pla 2010). (Assignatura optativa).
Curs: 2024 **Crèdits ECTS:** 6.0 **Idiomes:** Català, Castellà

PROFESSORAT

Professorat responsable: OSCAR ROMERO MORAL
Altres: Primer quadrimestre:
MARC MAYNOU YELAMOS - 11, 12
ANNA QUERALT CALAFAT - 11, 12
OSCAR ROMERO MORAL - 11, 12

CAPACITATS PRÈVIES

Es pressuposen coneixements en disseny físic de sistemes centralitzats, específicament en transaccions ACID i optimització física de consultes.

Ser capaç de llegir i entendre materials en anglès.

REQUISITS

- Pre-requisit DBD

COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

Específiques:

CES1.2. Donar solució a problemes d'integració en funció de les estratègies, dels estàndards i de les tecnologies disponibles.

CES1.5. Especificar, dissenyar, implementar i avaluar bases de dades.

CES1.6. Administrar bases de dades (CIS4.3)

CES3.2. Dissenyar i gestionar un magatzem de dades (data warehouse).

CT2.4. Demostrar coneixement i capacitat per a aplicar les eines necessàries a l'emmagatzematge, el processament i l'accés als sistemes d'informació, fins i tot els que es basen en la web.

Genèriques:

G8. Tenir motivació per a la realització professional i per a afrontar nous reptes, tenir una visió àmplia de les possibilitats de la carrera professional en l'àmbit de l'enginyeria en informàtica. Sentir-se motivat per la qualitat i la millora contínua, i actuar amb rigor en el desenvolupament professional. Capacitat d'adaptació als canvis organitzatius o tecnològics. Capacitat de treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

METODOLOGIES DOCENTS

A les hores de teoria el professor exposa els conceptes corresponents a algun dels continguts. Vàries de les sessions de teoria disposen de material addicional que l'alumne haurà de llegir abans de la sessió. Els dubtes que puguin sorgir al llegir aquest material es resoldran per part del professor o pels propis companys, mitjançant activitats d'aprenentatge cooperatiu, a les sessions de teoria. Alguns conceptes d'alguns continguts no són discutits pel professor sinó que els estudiants han de treballar materials que el professor haurà publicat al campus virtual (en aquest cas, caldrà resoldre petits qüestionaris que es trobaran al campus virtual).

Els laboratoris es treballen en parelles i es presenten durant una setmana i s'entreguen a la següent. A la primera sessió el professor introdueix el tema i l'objectiu de la pràctica. Els estudiants comencen a fer la pràctica i resolen els seus dubtes amb el professor. A la segona sessió es fa l'entrega.

El curs també té una component d'auto-aprenentatge, donat que hauran de treballar amb gestors no relacionals. A part del material de suport, han de ser capaços de resoldre dubtes o problemes d'us amb aquests gestors. S'activarà un forum per tal de promoure la discussió entre alumnes.

OBJECTIUS D'APRENENTATGE DE L'ASSIGNATURA

1. Enumerar, discutir i justificar amb detall els colls d'ampolla dels gestors relacionals enfront de sistemes d'emmagatzematge massius (de dades o processament).
2. Enumerar els pros i contres de tenir / no tenir un model únic de referència que s'adapti a tots els possibles escenaris d'emmagatzemament.
3. Enumerar, discutir i justificar amb detall els principis que comparteixen els nous sistemes d'emmagatzematge no relacionals.
4. Donat un escenari concret, amb uns requeriments d'usuari (parcials o totals), identificar quines característiques dels gestors relacionals actuarien com a coll d'ampolla (si cap) i enraonar quin gestor (o gestors) d'emmagatzematge alternatiu als relacionals és més adient.
5. Especificar i dissenyar arquitectures no relacionals que satisfacin els requeriments de qualsevol escenari.
6. Mantenir i evolucionar arquitectures no relacionals per adaptar-les a l'evolució dels requeriments d'usuari.
7. Donats uns sistemes d'emmagatzemament ja existents, analitzar, justificar i classificar els (possibles) problemes d'integració.
8. Especificar, dissenyar i crear processos de migració de dades (o processos ETL).
9. Donat un nivell de qualitat de dades, portar a terme migracions de dades entre sistemes que garanteixin el nivell de qualitat demanat.
10. Capacitat de treball en grup.
11. Capacitat de treball en situacions de falta d'informació.

HORES TOTALES DE DEDICACIÓ DE L'ESTUDIANTAT

Tipus	Hores	Percentatge
Hores grup petit	30,0	20.00
Hores grup gran	30,0	20.00
Hores activitats dirigides	6,0	4.00
Hores aprenentatge autònom	84,0	56.00

Dedicació total: 150 h

CONTINGUTS

Introducció

Descripció:

Paradigma one size fits all. Colls d'ampolla de les bases de dades relacionals. Els sistemes distribuïts no relacionals con a solució. Nous reptes i oportunitats.

Fonaments de bases de dades distribuïdes

Descripció:

Les noves arquitectures de bases de dades tenen un tret en comú: la distribució de les dades com a estratègia per afrontar volums ingents de dades. Aquest bloc es divideix en quatre grans apartats:

- 1) Introducció: Taxonomia de bases de dades distribuïdes. Arquitectures. Disseny de bases de dades distribuïdes (fragmentació i replicació). Mesures d'escalabilitat.
- 2) Optimització de consultes en sistemes distribuïts: Optimització semàntica, sintàctica i física de consultes. Paral·lelisme. Bases de dades paral·leles.
- 3) Concurrencia en sistemes distribuïts: Extensions del mòdul de concurrencia d'un sistema centralitzat per adaptar-ho a entorns distribuïts. CAP theorem. Two-phase locking. Multi-versió. Time-stamping. Tècniques optimistes. Eventually consistent.
- 4) Recuperació en sistemes distribuïts: Extensions del mòdul de recuperació d'un sistema centralitzat per adaptar-ho a entorns distribuïts. Checkpointing. Multi-level recovering. ARIES. 2-phase commit protocol. 3-phase commit protocol.

NOSQL

Descripció:

Sota l'etiqueta NOSQL trobem els nous sistemes distribuïts que relaxen tots (o alguns) dels colls d'ampolla del model relacional en entorns distribuïts. En aquest curs presentem una classificació d'aquests sistemes segons quina (o quines) característiques del model relacional relaxen i en veiem els 5 més rellevants:

- 1) Object-oriented databases. Introducció i assumpcions. Model de dades. Escalabilitat. Llenguatges d'alt nivell. La primera extensió del model relacional: object-relational databases. Similituds i diferències.
- 2) Column-oriented stores. Introducció i assumpcions. Model de dades. Escalabilitat. Llenguatges d'alt nivell. Extensions relacionals DSS.
- 3) Key-value stores. Introducció i assumpcions. Model de dades. Escalabilitat. Llenguatges d'alt nivell. MapReduce.
- 4) Document-stores. Introducció i assumpcions. Model de dades. XML databases. Escalabilitat. Llenguatges d'alt nivell.
- 5) Graph databases. Introducció i assumpcions. Model de dades. RDF i triple stores. Escalabilitat. Llenguatges d'alt nivell.

Integració de dades

Descripció:

El trencament del paradigma one size fits all comporta la co-existència de diferents gestors de bases de dades, cadascun especialitzat en un tipus d'escenari (inclús d'escenaris dintre del mateix sistema) amb unes característiques concretes. Integrar (per creuar, analitzar, homogenitzar, etc.) dades provinents de diferents sistemes és, avui dia, un repte. En aquest bloc ens centrem en com resoldre la integració de dades a nivell d'esquema:

- 1) Introducció: Tipus d'heterogeneïtats. Arquitectures: LAV, GAV, GLAV, Mediators, Peer-to-peer, Ontology-based data access.
- 2) Processos ETL: Originalment associats als magatzems de dades, avui dia tenen un àmbit més general i són sinònims de migració de dades entre sistemes (normalment heterogenis). El punt anterior afronta l'heterogeneïtat a nivell d'esquema, mentre que els processos ETL ho fan a nivell d'instància. Conceptes bàsics. Extracció, transformació i càrrega. Operadors ETL.
- 3) Molt lligat amb la integració de dades, diferents sistemes requereixen diferent nivell de qualitat de les dades, i això impacta directament en el nombre de transformacions a aplicar a les dades durant els processos de migració. Introducció. Causes. Mètodes i tècniques. Avaluació de la qualitat.

ACTIVITATS

Sessió 1: Motivació i introducció als sistemes NOSQL (més enllà de la tecnologia relacional).

Descripció:

Ve a classe i escolta. S'espera una actitud activa per part de l'alumne, ja sigui preguntant sobre els materials disponibles del tema, així com preparant i participant activament en les activitats d'aprenentatge actiu que s'aniran introduïnt durant el curs.

Objectius específics:

1, 2, 3

Dedicació: 2h

Grup gran/Teoria: 2h

Sessió 2: Fonaments dels sistemes de bases de dades distribuïts

Descripció:

Ve a classe i escolta. S'espera una actitud activa per part de l'alumne, ja sigui preguntant sobre els materials disponibles del tema, així com preparant i participant activament en les activitats d'aprenentatge actiu que s'aniran introduïnt durant el curs.

Objectius específics:

3

Dedicació: 4h

Aprenentatge autònom: 2h

Grup gran/Teoria: 2h

Sessió 2.1: Optimització de consultes en sistemes distribuïts

Descripció:

Ve a classe i escolta. S'espera una actitud activa per part de l'alumne, ja sigui preguntant sobre els materials disponibles del tema, així com preparant i participant activament en les activitats d'aprenentatge actiu que s'aniran introduïnt durant el curs.

Objectius específics:

1, 3

Dedicació: 4h

Aprenentatge autònom: 2h

Grup gran/Teoria: 2h

Sessió 2.2: Concurrència en sistemes distribuïts

Descripció:

Ve a classe i escolta. S'espera una actitud activa per part de l'alumne, ja sigui preguntant sobre els materials disponibles del tema, així com preparant i participant activament en les activitats d'aprenentatge actiu que s'aniran introduïnt durant el curs.

Objectius específics:

1, 3, 4

Dedicació: 4h

Aprenentatge autònom: 2h

Grup gran/Teoria: 2h

Sessió 2.3: Recuperació en sistemes distribuïts

Descripció:

Ve a classe i escolta. S'espera una actitud activa per part de l'alumne, ja sigui preguntant sobre els materials disponibles del tema, així com preparant i participant activament en les activitats d'aprenentatge actiu que s'aniran introduïnt durant el curs.

Objectius específics:

1, 3, 4

Dedicació: 4h

Aprenentatge autònom: 2h

Grup gran/Teoria: 2h

Sessió 3.1: Key-value stores

Descripció:

Ve a classe i escolta. S'espera una actitud activa per part de l'alumne, ja sigui preguntant sobre els materials disponibles del tema, així com preparant i participant activament en les activitats d'aprenentatge actiu que s'aniran introduïnt durant el curs.

Objectius específics:

4, 5, 6

Dedicació: 2h

Grup gran/Teoria: 2h

Sessió 3.2: MapReduce

Descripció:

Ve a classe i escolta. S'espera una actitud activa per part de l'alumne, ja sigui preguntant sobre els materials disponibles del tema, així com preparant i participant activament en les activitats d'aprenentatge actiu que s'aniran introduïnt durant el curs.

Objectius específics:

4, 6, 7, 8

Dedicació: 2h

Grup gran/Teoria: 2h

Sessió 3.3: Document-stores

Descripció:

Ve a classe i escolta. S'espera una actitud activa per part de l'alumne, ja sigui preguntant sobre els materials disponibles del tema, així com preparant i participant activament en les activitats d'aprenentatge actiu que s'aniran introduïnt durant el curs.

Objectius específics:

4, 5, 6

Dedicació: 2h

Grup gran/Teoria: 2h

Sessió 3.4: Object-relational

Descripció:

Ve a classe i escolta. S'espera una actitud activa per part de l'alumne, ja sigui preguntant sobre els materials disponibles del tema, així com preparant i participant activament en les activitats d'aprenentatge actiu que s'aniran introduïnt durant el curs.

Objectius específics:

2

Dedicació: 3h

Aprenentatge autònom: 1h

Grup gran/Teoria: 2h

Sessió 3.5: Graph databases

Descripció:

Ve a classe i escolta. S'espera una actitud activa per part de l'alumne, ja sigui preguntant sobre els materials disponibles del tema, així com preparant i participant activament en les activitats d'aprenentatge actiu que s'aniran introduïnt durant el curs.

Objectius específics:

4, 5, 6

Dedicació: 2h

Grup gran/Teoria: 2h

Sessió 3.6: Column-store databases

Descripció:

Ve a classe i escolta. S'espera una actitud activa per part de l'alumne, ja sigui preguntant sobre els materials disponibles del tema, així com preparant i participant activament en les activitats d'aprenentatge actiu que s'aniran introduïnt durant el curs.

Objectius específics:

4, 5, 6

Dedicació: 2h

Grup gran/Teoria: 2h

Sessió 4: Integració de dades

Descripció:

Ve a classe i escolta. S'espera una actitud activa per part de l'alumne, ja sigui preguntant sobre els materials disponibles del tema, així com preparant i participant activament en les activitats d'aprenentatge actiu que s'aniran introduïnt durant el curs.

Objectius específics:

2, 7

Dedicació: 2h

Grup gran/Teoria: 2h

Sessió 4.1: Qualitat de les dades

Descripció:

Ve a classe i escolta. S'espera una actitud activa per part de l'alumne, ja sigui preguntant sobre els materials disponibles del tema, així com preparant i participant activament en les activitats d'aprenentatge actiu que s'aniran introduïnt durant el curs.

Objectius específics:

2, 8, 9

Dedicació: 2h

Grup gran/Teoria: 2h

Sessió 4.2: Processos ETL

Descripció:

Ve a classe i escolta. S'espera una actitud activa per part de l'alumne, ja sigui preguntant sobre els materials disponibles del tema, així com preparant i participant activament en les activitats d'aprenentatge actiu que s'aniran introduïnt durant el curs.

Objectius específics:

2, 8

Dedicació: 2h

Grup gran/Teoria: 2h

Lab 1: Pràctica sobre la teoria dels sistemes distribuïts

Objectius específics:

1, 3, 10, 11

Competències relacionades:

G8. Tenir motivació per a la realització professional i per a afrontar nous reptes, tenir una visió àmplia de les possibilitats de la carrera professional en l'àmbit de l'enginyeria en informàtica. Sentir-se motivat per la qualitat i la millora contínua, i actuar amb rigor en el desenvolupament professional. Capacitat d'adaptació als canvis organitzatius o tecnològics. Capacitat de treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

Dedicació: 8h

Aprenentatge autònom: 5h

Activitats dirigides: 3h

Lab: Key-value stores (part 1)

Objectius específics:

4, 5, 6, 10, 11

Competències relacionades:

G8. Tenir motivació per a la realització professional i per a afrontar nous reptes, tenir una visió àmplia de les possibilitats de la carrera professional en l'àmbit de l'enginyeria en informàtica. Sentir-se motivat per la qualitat i la millora contínua, i actuar amb rigor en el desenvolupament professional. Capacitat d'adaptació als canvis organitzatius o tecnològics. Capacitat de treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

Dedicació: 13h

Aprenentatge autònom: 10h

Activitats dirigides: 3h

Lab: Key-value stores (part 2)

Objectius específics:

4, 5, 6, 10, 11

Competències relacionades:

G8. Tenir motivació per a la realització professional i per a afrontar nous reptes, tenir una visió àmplia de les possibilitats de la carrera professional en l'àmbit de l'enginyeria en informàtica. Sentir-se motivat per la qualitat i la millora contínua, i actuar amb rigor en el desenvolupament professional. Capacitat d'adaptació als canvis organitzatius o tecnològics. Capacitat de treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

Dedicació: 11h

Aprenentatge autònom: 8h

Activitats dirigides: 3h

Lab: Document-stores (part 1)

Objectius específics:

4, 5, 6, 10, 11

Competències relacionades:

G8. Tenir motivació per a la realització professional i per a afrontar nous reptes, tenir una visió àmplia de les possibilitats de la carrera professional en l'àmbit de l'enginyeria en informàtica. Sentir-se motivat per la qualitat i la millora contínua, i actuar amb rigor en el desenvolupament professional. Capacitat d'adaptació als canvis organitzatius o tecnològics. Capacitat de treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

Dedicació: 13h

Aprenentatge autònom: 10h

Activitats dirigides: 3h

Lab: Document-stores (part 2)

Objectius específics:

4, 5, 6, 10, 11

Competències relacionades:

G8. Tenir motivació per a la realització professional i per a afrontar nous reptes, tenir una visió àmplia de les possibilitats de la carrera professional en l'àmbit de l'enginyeria en informàtica. Sentir-se motivat per la qualitat i la millora contínua, i actuar amb rigor en el desenvolupament professional. Capacitat d'adaptació als canvis organitzatius o tecnològics. Capacitat de treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

Dedicació: 11h

Aprenentatge autònom: 8h

Activitats dirigides: 3h

Lab: Object-relational Vs. Object-relational databases (part 1)

Objectius específics:

2, 4, 10, 11

Competències relacionades:

G8. Tenir motivació per a la realització professional i per a afrontar nous reptes, tenir una visió àmplia de les possibilitats de la carrera professional en l'àmbit de l'enginyeria en informàtica. Sentir-se motivat per la qualitat i la millora contínua, i actuar amb rigor en el desenvolupament professional. Capacitat d'adaptació als canvis organitzatius o tecnològics. Capacitat de treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

Dedicació: 7h

Aprenentatge autònom: 4h

Activitats dirigides: 3h

Lab: Object-relational Vs. Object-relational databases (part 2)

Objectius específics:

2, 4, 5, 6, 10, 11

Competències relacionades:

G8. Tenir motivació per a la realització professional i per a afrontar nous reptes, tenir una visió àmplia de les possibilitats de la carrera professional en l'àmbit de l'enginyeria en informàtica. Sentir-se motivat per la qualitat i la millora contínua, i actuar amb rigor en el desenvolupament professional. Capacitat d'adaptació als canvis organitzatius o tecnològics. Capacitat de treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

Dedicació: 7h

Aprenentatge autònom: 4h

Activitats dirigides: 3h

Lab: Graph databases (part 1)

Objectius específics:

4, 5, 6, 10, 11

Competències relacionades:

G8. Tenir motivació per a la realització professional i per a afrontar nous reptes, tenir una visió àmplia de les possibilitats de la carrera professional en l'àmbit de l'enginyeria en informàtica. Sentir-se motivat per la qualitat i la millora contínua, i actuar amb rigor en el desenvolupament professional. Capacitat d'adaptació als canvis organitzatius o tecnològics. Capacitat de treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

Dedicació: 9h

Aprenentatge autònom: 6h

Activitats dirigides: 3h

Lab: Graph databases (part 2)

Objectius específics:

4, 5, 6, 10, 11

Competències relacionades:

G8. Tenir motivació per a la realització professional i per a afrontar nous reptes, tenir una visió àmplia de les possibilitats de la carrera professional en l'àmbit de l'enginyeria en informàtica. Sentir-se motivat per la qualitat i la millora contínua, i actuar amb rigor en el desenvolupament professional. Capacitat d'adaptació als canvis organitzatius o tecnològics. Capacitat de treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

Dedicació: 13h

Aprenentatge autònom: 10h

Activitats dirigides: 3h

Lab 6: Simulació de column-oriented DBs

Objectius específics:

4, 5, 6, 10, 11

Competències relacionades:

G8. Tenir motivació per a la realització professional i per a afrontar nous reptes, tenir una visió àmplia de les possibilitats de la carrera professional en l'àmbit de l'enginyeria en informàtica. Sentir-se motivat per la qualitat i la millora contínua, i actuar amb rigor en el desenvolupament professional. Capacitat d'adaptació als canvis organitzatius o tecnològics. Capacitat de treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

Dedicació: 13h

Aprenentatge autònom: 10h

Activitats dirigides: 3h

Exàmen final

Descripció:

Exàmen final de tots els continguts de l'assignatura

Objectius específics:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11

Competències relacionades:

G8. Tenir motivació per a la realització professional i per a afrontar nous reptes, tenir una visió àmplia de les possibilitats de la carrera professional en l'àmbit de l'enginyeria en informàtica. Sentir-se motivat per la qualitat i la millora contínua, i actuar amb rigor en el desenvolupament professional. Capacitat d'adaptació als canvis organitzatius o tecnològics. Capacitat de treballar en situacions de carència d'informació i/o amb restriccions temporals i/o de recursos.

Dedicació: 8h

Aprenentatge autònom: 6h

Activitats dirigides: 2h

SISTEMA DE QUALIFICACIÓ

Nota final = 70% min(10,P) + 20%EX + 10%C

EX = nota de l'examen final

P = promig ponderat de les notes dels qüestionaris i de les sessions de laboratori

C = avaluació entre companys

Càlcul de P:

1) Multiplicar la nota obtinguda a cadascuna de les activitats per un pes igual a 1, 2, 4 o 8 (depenent del contingut de l'activitat en qüestió)

2) Dividir la suma d'aquests valors per la suma de pesos assignats menys el 10%

Càlcul de C: els estudiants tindran diverses parelles durant el semestre. Al final hauran de valorar-les segons 5 criteris: contribucions, abilitat per resoldre problemes, actitud, preparació, treball en equip i concentració. Prenent com a base aquestes valoracions, el professor assignarà la nota.

La qualificació de la competència transversal "Actitud adequada davant el treball" serà A (competència superada amb excel·lència), B (competència superada al nivell desitjat), C (competència superada a un nivell suficient) o D (competència no superada). Aquesta nota es computa com la mitja ponderada de C i la nota d'actitud recollida pel professor (participació en les activitats dirigides a classe i grau d'autonomia a l'hora de preparar les sessions NOSQL).

BIBLIOGRAFIA

Bàsica:

- Garcia-Molina, Hector; Ullman, Jeffrey D; Widom, Jennifer. Database systems: the complete book. Second edition, Pearson new international edition. Harlow, Essex: Pearson Education Limited, [2014]. ISBN 9781292024479.
- Ozsu, M.T.; Valduriez, P. Principles of distributed database systems. 4th ed. New York: Springer, 2020. ISBN 9783030262525.
- Sadalage, P.J.; Fowler, M. NoSQL distilled: a brief guide to the emerging world of polygot persistence. Boston, Mass. ; London: Addison-Wesley, 2013. ISBN 978-0321826626.
- Abiteboul, S. Web data management. Cambridge University Press, 2012. ISBN 978-1107012431.

Complementària:

- Liu, L.; Özsu, Tamer, M. (eds.). Encyclopedia of database systems. New York ; London: Springer, 2009. ISBN 9780387399409.
- Plattner, H.; Zeier, A. In-memory data management: technology and applications. 2nd ed. Springer, 2012. ISBN 978-3642295744.