



Course guide

270087 - CBDE - Concepts for Specialised Databases

Last modified: 13/07/2023

Unit in charge: Barcelona School of Informatics
Teaching unit: 747 - ESSI - Department of Service and Information System Engineering.

Degree: BACHELOR'S DEGREE IN INFORMATICS ENGINEERING (Syllabus 2010). (Optional subject).

Academic year: 2023 **ECTS Credits:** 6.0 **Languages:** Catalan, Spanish

LECTURER

Coordinating lecturer: OSCAR ROMERO MORAL

Others:
Primer quadrimestre:
SERGI NADAL FRANCESCH - 11, 12
ANNA QUERALT CALAFAT - 11, 12
OSCAR ROMERO MORAL - 11, 12

PRIOR SKILLS

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.

REQUIREMENTS

- Prerequisite DBD

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:

CES1.2. To solve integration problems in function of the strategies, standards and available technologies

CES1.5. To specify, design, implement and evaluate databases.

CES1.6. To administrate databases (CIS4.3).

CES3.2. To design and manage a data warehouse.

CT2.4. To demonstrate knowledge and capacity to apply the needed tools for storage, processing and access to the information system, even if they are web-based systems.

General:

G8. APPROPRIATE ATTITUDE TOWARDS WORK: to have motivation to be professional and to face new challenges, have a width vision of the possibilities of the career in the field of informatics engineering. To feel motivated for the quality and the continuous improvement, and behave rigorously in the professional development. Capacity to adapt oneself to organizational or technological changes. Capacity to work in situations with information shortage and/or time and/or resources restrictions.



TEACHING METHODOLOGY

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 introduceix 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.

LEARNING OBJECTIVES OF THE SUBJECT

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.

STUDY LOAD

Type	Hours	Percentage
Self study	84,0	56.00
Hours large group	30,0	20.00
Hours small group	30,0	20.00
Guided activities	6,0	4.00

Total learning time: 150 h

CONTENTS

Introducció

Description:

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

Description:

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) Concurrència en sistemes distribuïts: Extensions del mòdul de concurrència 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

Description:

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 assumptions. 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 assumptions. Model de dades. Escalabilitat. Llenguatges d'alt nivell. Extensions relacionals DSS.
- 3) Key-value stores. Introducció i assumptions. Model de dades. Escalabilitat. Llenguatges d'alt nivell. MapReduce.
- 4) Document-stores. Introducció i assumptions. Model de dades. XML databases. Escalabilitat. Llenguatges d'alt nivell.
- 5) Graph databases. Introducció i assumptions. Model de dades. RDF i triple stores. Escalabilitat. Llenguatges d'alt nivell.

Integració de dades

Description:

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 (inclus 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.



ACTIVITIES

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

Description:

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 introduint durant el curs.

Specific objectives:

1, 2, 3

Full-or-part-time: 2h

Theory classes: 2h

Sessió 2: Fonaments dels sistemes de bases de dades distribuïts

Description:

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 introduint durant el curs.

Specific objectives:

3

Full-or-part-time: 4h

Theory classes: 2h

Self study: 2h

Sessió 2.1: Optimització de consultes en sistemes distribuïts

Description:

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 introduint durant el curs.

Specific objectives:

1, 3

Full-or-part-time: 4h

Theory classes: 2h

Self study: 2h

Sessió 2.2: Concurrència en sistemes distribuïts

Description:

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 introduint durant el curs.

Specific objectives:

1, 3, 4

Full-or-part-time: 4h

Theory classes: 2h

Self study: 2h



Sessió 2.3: Recuperació en sistemes distribuïts

Description:

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 introduint durant el curs.

Specific objectives:

1, 3, 4

Full-or-part-time: 4h

Theory classes: 2h

Self study: 2h

Sessió 3.1: Key-value stores

Description:

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 introduint durant el curs.

Specific objectives:

4, 5, 6

Full-or-part-time: 2h

Theory classes: 2h

Sessió 3.2: MapReduce

Description:

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 introduint durant el curs.

Specific objectives:

4, 6, 7, 8

Full-or-part-time: 2h

Theory classes: 2h

Sessió 3.3: Document-stores

Description:

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 introduint durant el curs.

Specific objectives:

4, 5, 6

Full-or-part-time: 2h

Theory classes: 2h



Sessió 3.4: Object-relational

Description:

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 introduint durant el curs.

Specific objectives:

2

Full-or-part-time: 3h

Theory classes: 2h

Self study: 1h

Sessió 3.5: Graph databases

Description:

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 introduint durant el curs.

Specific objectives:

4, 5, 6

Full-or-part-time: 2h

Theory classes: 2h

Sessió 3.6: Column-store databases

Description:

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 introduint durant el curs.

Specific objectives:

4, 5, 6

Full-or-part-time: 2h

Theory classes: 2h

Sessió 4: Integració de dades

Description:

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 introduint durant el curs.

Specific objectives:

2, 7

Full-or-part-time: 2h

Theory classes: 2h



Sessió 4.1: Qualitat de les dades

Description:

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 introduint durant el curs.

Specific objectives:

2, 8, 9

Full-or-part-time: 2h

Theory classes: 2h

Sessió 4.2: Processos ETL

Description:

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 introduint durant el curs.

Specific objectives:

2, 8

Full-or-part-time: 2h

Theory classes: 2h

Lab 1: Pràctica sobre la teoria dels sistemes distribuïts

Specific objectives:

1, 3, 10, 11

Related competencies :

G8. APPROPRIATE ATTITUDE TOWARDS WORK: to have motivation to be professional and to face new challenges, have a width vision of the possibilities of the career in the field of informatics engineering. To feel motivated for the quality and the continuous improvement, and behave rigorously in the professional development. Capacity to adapt oneself to organizational or technological changes. Capacity to work in situations with information shortage and/or time and/or resources restrictions.

Full-or-part-time: 8h

Guided activities: 3h

Self study: 5h

Lab: Key-value stores (part 1)

Specific objectives:

4, 5, 6, 10, 11

Related competencies :

G8. APPROPRIATE ATTITUDE TOWARDS WORK: to have motivation to be professional and to face new challenges, have a width vision of the possibilities of the career in the field of informatics engineering. To feel motivated for the quality and the continuous improvement, and behave rigorously in the professional development. Capacity to adapt oneself to organizational or technological changes. Capacity to work in situations with information shortage and/or time and/or resources restrictions.

Full-or-part-time: 13h

Guided activities: 3h

Self study: 10h



Lab: Key-value stores (part 2)

Specific objectives:

4, 5, 6, 10, 11

Related competencies :

G8. APPROPRIATE ATTITUDE TOWARDS WORK: to have motivation to be professional and to face new challenges, have a width vision of the possibilities of the career in the field of informatics engineering. To feel motivated for the quality and the continuous improvement, and behave rigorously in the professional development. Capacity to adapt oneself to organizational or technological changes. Capacity to work in situations with information shortage and/or time and/or resources restrictions.

Full-or-part-time: 11h

Guided activities: 3h

Self study: 8h

Lab: Document-stores (part 1)

Specific objectives:

4, 5, 6, 10, 11

Related competencies :

G8. APPROPRIATE ATTITUDE TOWARDS WORK: to have motivation to be professional and to face new challenges, have a width vision of the possibilities of the career in the field of informatics engineering. To feel motivated for the quality and the continuous improvement, and behave rigorously in the professional development. Capacity to adapt oneself to organizational or technological changes. Capacity to work in situations with information shortage and/or time and/or resources restrictions.

Full-or-part-time: 13h

Guided activities: 3h

Self study: 10h

Lab: Document-stores (part 2)

Specific objectives:

4, 5, 6, 10, 11

Related competencies :

G8. APPROPRIATE ATTITUDE TOWARDS WORK: to have motivation to be professional and to face new challenges, have a width vision of the possibilities of the career in the field of informatics engineering. To feel motivated for the quality and the continuous improvement, and behave rigorously in the professional development. Capacity to adapt oneself to organizational or technological changes. Capacity to work in situations with information shortage and/or time and/or resources restrictions.

Full-or-part-time: 11h

Guided activities: 3h

Self study: 8h



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

Specific objectives:

2, 4, 10, 11

Related competencies :

G8. APPROPRIATE ATTITUDE TOWARDS WORK: to have motivation to be professional and to face new challenges, have a width vision of the possibilities of the career in the field of informatics engineering. To feel motivated for the quality and the continuous improvement, and behave rigorously in the professional development. Capacity to adapt oneself to organizational or technological changes. Capacity to work in situations with information shortage and/or time and/or resources restrictions.

Full-or-part-time: 7h

Guided activities: 3h

Self study: 4h

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

Specific objectives:

2, 4, 5, 6, 10, 11

Related competencies :

G8. APPROPRIATE ATTITUDE TOWARDS WORK: to have motivation to be professional and to face new challenges, have a width vision of the possibilities of the career in the field of informatics engineering. To feel motivated for the quality and the continuous improvement, and behave rigorously in the professional development. Capacity to adapt oneself to organizational or technological changes. Capacity to work in situations with information shortage and/or time and/or resources restrictions.

Full-or-part-time: 7h

Guided activities: 3h

Self study: 4h

Lab: Graph databases (part 1)

Specific objectives:

4, 5, 6, 10, 11

Related competencies :

G8. APPROPRIATE ATTITUDE TOWARDS WORK: to have motivation to be professional and to face new challenges, have a width vision of the possibilities of the career in the field of informatics engineering. To feel motivated for the quality and the continuous improvement, and behave rigorously in the professional development. Capacity to adapt oneself to organizational or technological changes. Capacity to work in situations with information shortage and/or time and/or resources restrictions.

Full-or-part-time: 9h

Guided activities: 3h

Self study: 6h



Lab: Graph databases (part 2)

Specific objectives:

4, 5, 6, 10, 11

Related competencies :

G8. APPROPRIATE ATTITUDE TOWARDS WORK: to have motivation to be professional and to face new challenges, have a width vision of the possibilities of the career in the field of informatics engineering. To feel motivated for the quality and the continuous improvement, and behave rigorously in the professional development. Capacity to adapt oneself to organizational or technological changes. Capacity to work in situations with information shortage and/or time and/or resources restrictions.

Full-or-part-time: 13h

Guided activities: 3h

Self study: 10h

Lab 6: Simulacio de column-oriented DBs

Specific objectives:

4, 5, 6, 10, 11

Related competencies :

G8. APPROPRIATE ATTITUDE TOWARDS WORK: to have motivation to be professional and to face new challenges, have a width vision of the possibilities of the career in the field of informatics engineering. To feel motivated for the quality and the continuous improvement, and behave rigorously in the professional development. Capacity to adapt oneself to organizational or technological changes. Capacity to work in situations with information shortage and/or time and/or resources restrictions.

Full-or-part-time: 13h

Guided activities: 3h

Self study: 10h

Final exam

Description:

Final exam including all the course content

Specific objectives:

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

Related competencies :

G8. APPROPRIATE ATTITUDE TOWARDS WORK: to have motivation to be professional and to face new challenges, have a width vision of the possibilities of the career in the field of informatics engineering. To feel motivated for the quality and the continuous improvement, and behave rigorously in the professional development. Capacity to adapt oneself to organizational or technological changes. Capacity to work in situations with information shortage and/or time and/or resources restrictions.

Full-or-part-time: 8h

Guided activities: 2h

Self study: 6h



GRADING SYSTEM

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 = evaluació entre companyys

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. Prencent 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).

BIBLIOGRAPHY

Basic:

- 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.
- Ozsü, 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.

Complementary:

- Liu, L.; Özsü, Tamer, M. (eds.). Encyclopedia of database systems [on line]. New York ; London: Springer, 2009 [Consultation: 27/07/2023]. Available on: <https://link.springer.com/recursos.biblioteca.upc.edu/referencework/10.1007/978-0-387-39940-9>. ISBN 9780387399409.
- Plattner, H.; Zeier, A. In-memory data management: technology and applications. 2nd ed. Springer, 2012. ISBN 978-3642295744.