



Guia docent

270221 - BDA - Bases de Dades Avançades

Última modificació: 19/07/2023

Unitat responsable: Facultat d'Informàtica de Barcelona
Unitat que imparteix: 747 - ESSI - Departament d'Enginyeria de Serveis i Sistemes d'Informació.
Titulació: GRAU EN CIÈNCIA I ENGINYERIA DE DADES (Pla 2017). (Assignatura obligatòria).
Curs: 2023 **Crèdits ECTS:** 6.0 **Idiomes:** Català, Anglès

PROFESSORAT

Professorat responsable: ALBERTO ABELLO GAMAZO
Altres: Primer quadrimestre:
ALBERTO ABELLO GAMAZO - 11, 12, 13
BESIM BILALLI - 11, 12, 13

CAPACITATS PRÈVIES

Ser capaç de llegir i entendre materials en anglès.
Ser capaç d'enumerar les etapes que conformen el procés de l'enginyeria de software.
Ser capaç d'entendre esquemes conceptuals en UML.
Ser capaç de crear, consultar i manipular bases de dades amb SQL.

COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

Específiques:

CE7. Demostrar coneixement i capacitat d'aplicació de les eines necessàries per a l'emmagatzematge, el processament i l'accés a les dades.

Genèriques:

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.
CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.

Transversals:

CT4. Treball en equip. Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, assumint compromisos tenint en compte els recursos disponibles.
CT6. Aprenentatge autònom. Detectar deficiències en el propi coneixement i superar-les mitjançant la reflexió crítica i l'elecció de la millor actuació per ampliar aquest coneixement.

Bàsiques:

CB2. Que els estudiants sàpiguen aplicar els seus coneixements al seu treball o vocació d'una manera professional i posseeixin les competències que solen demostrar-se mitjançant l'elaboració i defensa d'arguments i la resolució de problemes dins la seva àrea d'estudi.
CB3. Que els estudiants tinguin la capacitat de reunir i interpretar dades rellevants (normalment dins la seva àrea d'estudi) per emetre judicis que incloguin una reflexió sobre temes rellevants d'índole social, científica o ètica.

METODOLOGIES DOCENTS

L'assignatura es compon de teoria, i sessions de laboratori.

Teoria: S'utilitzaran tècniques de classe invertida que requereixen que l'estudiant treballi materials multimedia abans de la classe. Les classes de teoria consisteixen en explicacions complementàries del professor i resolució de problemes.

Laboratori: S'utilitzaran eines representatives per a l'aplicació de conceptes teòrics (per exemple, Indyco Builder, PotgreSQL, Pentaho Data Integration, Spark). Es faran també dos projectes, en què els estudiants treballaran en equips: un d'anàlisi descriptiu de dades en un magatzem de dades i l'altre d'anàlisi predictiu en un entorn Big Data. Conseqüentment, hi haurà dos entregables fora de l'horari de classe, però els estudiants també seran avaluats individualment a l'aula sobre els coneixements adquirits durant cadascun dels projectes.

El curs té una component d'aprenentatge autònom, donat que els estudiants hauran de treballar amb diferents eines de gestió i processament de dades. A part del material de suport, els estudiants han de ser capaços de resoldre dubtes o problemes d'ús d'aquestes eines.

OBJECTIUS D'APRENTATGE DE L'ASSIGNATURA

1. Ser capaç de discutir i justificar amb detall els principis arquitectònics i colls d'ampolla dels gestors relacionals enfront de sistemes d'emmagatzematge i processament alternatius.
2. Ser capaç d'obtenir l'esquema lògic d'un magatzem de dades a partir d'un esquema conceptual expressat en UML, detectar i corregir els seus defectes.
3. Ser capaç de triar i justificar l'ús d'emmagatzemament basat en files o columnes.
4. Ser capaç d'explicar i utilitzar els principals mecanismes de processament paral·lel de consultes en entorns distribuïts, i detectar colls d'ampolla.
5. Ser capaç de justificar i utilitzar entorns de processament distribuït de dades d'estil funcional, tipus MapReduce/Spark.

HORES TOTALES DE DEDICACIÓ DE L'ESTUDIANTAT

Tipus	Hores	Percentatge
Hores grup gran	30,0	20.00
Hores aprenentatge autònom	90,0	60.00
Hores grup petit	30,0	20.00

Dedicació total: 150 h

CONTINGUTS

Introducció

Descripció:

Emmagatzemament de dades en entorns analítics i Big Data

Magatzems de dades

Descripció:

Emmagatzemament de dades. Fluxes de dades ETL. Integració de dades. Eines OLAP. Tècniques de compressió i emmagatzemament columnar.



Bases de dades distribuïdes

Descripció:

Taxonomia de bases de dades distribuïdes. Arquitectures. Disseny de bases de dades distribuïdes (fragmentació i replicació). Paral·lisme. Mesures d'escalabilitat. Sistemes de fitxers distribuïts.

Processament distribuït de dades

Descripció:

Importància de l'accés seqüencial paral·lel. Barreres de sincronització (Bulk Synchronous Parallel model). Entorns de processament distribuït de dades d'estil funcional (MapReduce i Spark). Abstracció de conjunts de dades distribuïts (Resilient Distributed Datasets). Arquitectures de Big Data.

ACTIVITATS

Introducció

Descripció:

Introducció de l'assignatura, motivació i visió general de les eines de gestió de dades existents actualment, els seus avantatges i inconvenients

Objectius específics:

1

Competències relacionades:

CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.

CE7. Demostrar coneixement i capacitat d'aplicació de les eines necessàries per a l'emmagatzematge, el processament i l'accés a les dades.

CT4. Treball en equip. Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, assumint compromisos tenint en compte els recursos disponibles.

CT6. Aprenentatge autònom. Detectar deficiències en el propi coneixement i superar-les mitjançant la reflexió crítica i l'elecció de la millor actuació per ampliar aquest coneixement.

CB2. Que els estudiants sàpiguen aplicar els seus coneixements al seu treball o vocació d'una manera professional i posseeixin les competències que solen demostrar-se mitjançant l'elaboració i defensa d'arguments i la resolució de problemes dins la seva àrea d'estudi.

CB3. Que els estudiants tinguin la capacitat de reunir i interpretar dades rellevants (normalment dins la seva àrea d'estudi) per emetre judicis que incloguin una reflexió sobre temes rellevants d'índole social, científica o ètica.

Dedicació: 2h

Grup gran/Teoria: 2h



Estudi dels magatzems de dades

Objectius específics:

2, 3

Competències relacionades:

CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.
CE7. Demostrar coneixement i capacitat d'aplicació de les eines necessàries per a l'emmagatzematge, el processament i l'accés a les dades.

CT4. Treball en equip. Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, assumint compromisos tenint en compte els recursos disponibles.

CT6. Aprenentatge autònom. Detectar deficiències en el propi coneixement i superar-les mitjançant la reflexió crítica i l'elecció de la millor actuació per ampliar aquest coneixement.

CB2. Que els estudiants sàpiguen aplicar els seus coneixements al seu treball o vocació d'una manera professional i posseeixin les competències que solen demostrar-se mitjançant l'elaboració i defensa d'arguments i la resolució de problemes dins la seva àrea d'estudi.

CB3. Que els estudiants tinguin la capacitat de reunir i interpretar dades rellevants (normalment dins la seva àrea d'estudi) per emetre judicis que incloguin una reflexió sobre temes rellevants d'índole social, científica o ètica.

Dedicació: 62h

Grup gran/Teoria: 10h

Grup petit/Laboratori: 14h

Aprenentatge autònom: 38h

Estudi de bases de dades distribuïdes

Descripció:

Aprenentatge dels principis de bases de dades distribuïdes i la seva aplicació en sistemes NOSQL

Objectius específics:

1, 4

Competències relacionades:

CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.

CE7. Demostrar coneixement i capacitat d'aplicació de les eines necessàries per a l'emmagatzematge, el processament i l'accés a les dades.

CT4. Treball en equip. Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, assumint compromisos tenint en compte els recursos disponibles.

CT6. Aprenentatge autònom. Detectar deficiències en el propi coneixement i superar-les mitjançant la reflexió crítica i l'elecció de la millor actuació per ampliar aquest coneixement.

CB2. Que els estudiants sàpiguen aplicar els seus coneixements al seu treball o vocació d'una manera professional i posseeixin les competències que solen demostrar-se mitjançant l'elaboració i defensa d'arguments i la resolució de problemes dins la seva àrea d'estudi.

CB3. Que els estudiants tinguin la capacitat de reunir i interpretar dades rellevants (normalment dins la seva àrea d'estudi) per emetre judicis que incloguin una reflexió sobre temes rellevants d'índole social, científica o ètica.

Dedicació: 14h

Grup gran/Teoria: 6h

Grup petit/Laboratori: 4h

Aprenentatge autònom: 4h



Estudi del processament distribuït de dades

Descripció:

Aprenentatge de tècniques de processament distribuït de dades en entorns d'estil funcional

Objectius específics:

1, 4, 5

Competències relacionades:

CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.

CE7. Demostrar coneixement i capacitat d'aplicació de les eines necessàries per a l'emmagatzematge, el processament i l'accés a les dades.

CT4. Treball en equip. Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, assumint compromisos tenint en compte els recursos disponibles.

CT6. Aprenentatge autònom. Detectar deficiències en el propi coneixement i superar-les mitjançant la reflexió crítica i l'elecció de la millor actuació per ampliar aquest coneixement.

CB2. Que els estudiants sàpiguen aplicar els seus coneixements al seu treball o vocació d'una manera professional i posseeixin les competències que solen demostrar-se mitjançant l'elaboració i defensa d'arguments i la resolució de problemes dins la seva àrea d'estudi.

CB3. Que els estudiants tinguin la capacitat de reunir i interpretar dades rellevants (normalment dins la seva àrea d'estudi) per emetre judicis que incloguin una reflexió sobre temes rellevants d'índole social, científica o ètica.

Dedicació: 60h

Grup gran/Teoria: 10h

Grup petit/Laboratori: 12h

Aprenentatge autònom: 38h



Examen final

Descripció:

Examen global de l'assignatura

Objectius específics:

1, 2, 3, 4, 5

Competències relacionades:

CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.

CE7. Demostrar coneixement i capacitat d'aplicació de les eines necessàries per a l'emmagatzematge, el processament i l'accés a les dades.

CT4. Treball en equip. Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, assumint compromisos tenint en compte els recursos disponibles.

CT6. Aprenentatge autònom. Detectar deficiències en el propi coneixement i superar-les mitjançant la reflexió crítica i l'elecció de la millor actuació per ampliar aquest coneixement.

CB2. Que els estudiants sàpiguen aplicar els seus coneixements al seu treball o vocació d'una manera professional i posseeixin les competències que solen demostrar-se mitjançant l'elaboració i defensa d'arguments i la resolució de problemes dins la seva àrea d'estudi.

CB3. Que els estudiants tinguin la capacitat de reunir i interpretar dades rellevants (normalment dins la seva àrea d'estudi) per emetre judicis que incloguin una reflexió sobre temes rellevants d'índole social, científica o ètica.

Dedicació: 12h

Activitats dirigides: 2h

Aprenentatge autònom: 10h

SISTEMA DE QUALIFICACIÓ

Nota final = $\min(10 ; \max(20\%EP+40\%EF ; 60\% EF) + 40\% P + 10\% C)$

EP = nota de l'examen parcial

EF = nota de l'examen final

P = nota de projecte, com a promig ponderat dels projectes del curs

C = participació a classe

Pels estudiants que puguin concórrer a la reavaluació, la nota de examen de reavaluació substituirà EF.



BIBLIOGRAFIA

Bàsica:

- Garcia-Molina, Hector; Ullman, Jeffrey D; Widom, Jennifer. Database systems : the complete book [en línia]. Second edition, Pearson new international edition. Essex: Pearson Education, 2013 [Consulta: 02/08/2023]. Disponible a: <https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=5174436>. ISBN 9781292024479.
- Database Technologies and Information Management. Slides on Advanced Databases course.
- Golfarelli, M.; Rizzi, S. Data warehouse design: modern principles and methodologies. New York [etc.]: McGraw Hill, 2009. ISBN 9780071610391.
- Vaisman, A.; Zimányi, E. Data warehouse systems: design and implementation. Second edition. Berlin: Springer, 2022. ISBN 9783662651667.
- Özsü, M.T.; Valdúriez, 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. Addison-Wesley, 2013. ISBN 9780321826626.
- Badia, Antonio. SQL for data science : data cleaning, wrangling and analytics with relational databases. Springer, 2020. ISBN 9783030575915.
- Abelló, Alberto; Jovanovic, Petar. Data Warehousing and OLAP.
- Abelló, Alberó; Nadal, Sergi. Big Data Management.

Complementària:

- Exercises Big Data Management.
- Exercises Data Warehousing.

RECURSOS

Enllaç web:

- <https://bdma.ulb.ac.be/bdma>. Erasmus Mundus Master on Big Data Management and Analytics
- <https://cs.ulb.ac.be/conferences/ebiss.html>. European Big Data Management and Analytics Summer School (eBISS)