



Guia docent

270952 - DW - Magatzems de Dades

Última modificació: 23/11/2023

Unitat responsable: Facultat d'Informàtica de Barcelona
Unitat que imparteix: 747 - ESSI - Departament d'Enginyeria de Serveis i Sistemes d'Informació.
Titulació: MÀSTER UNIVERSITARI EN CIÈNCIA DE DADES (Pla 2021). (Assignatura obligatòria).
Curs: 2023 **Crèdits ECTS:** 6.0 **Idiomes:** Anglès

PROFESSORAT

Professorat responsable: ALBERTO ABELLO GAMAZO
Altres: Primer quadrimestre:
ALBERTO ABELLO GAMAZO - 11, 12
PETAR JOVANOVIC - 11, 12

CAPACITATS PRÈVIES

Coneixement bàsic de bases de dades relacionals i SQL.

Específicament, s'assumirà coneixements de:

- Diagrames de classes UML
- Àlgebra relacional
- Consultes SQL
- Vistes relacionals
- Operacions d'arbres-B (insercions i splits)
- Conceptes bàsic d'optimització física de consultes

COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

Específiques:

- CE2. Aplicar els fonaments de la gestió i processament de dades en un problema de ciència de dades
- CE3. Aplicar mètodes d'integració de dades per donar solució a problemes de ciència de dades en entorns heterogenis
- CE5. Modelar, dissenyar i implementar sistemes complexos de dades, incloent-hi la visualització de dades
- CE7. Identificar les limitacions imposades per la qualitat de dades en un problema de ciència de dades i aplicar tècniques per a disminuir el seu impacte

Genèriques:

- CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

Transversals:

- CT1. EMPRENEDORIA I INNOVACIÓ: Conèixer i comprendre l'organització d'una empresa i les ciències que regeixen la seva activitat; tenir capacitat per entendre les normes laborals i les relacions entre la planificació, les estratègies industrials i comercials, la qualitat i el benefici. Conèixer i entendre els mecanismes en què es basa la recerca científica, així com els mecanismes i instruments de transferència de resultats entre els diferents agents socioeconòmics implicats en els processos d'I+D+i.
- CT3. TREBALL EN EQUIP: Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o duent a terme tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, tot assumint compromisos considerant els recursos disponibles.
- CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

Bàsiques:

CB10. Posseir i comprendre coneixements que aportin una base o oportunitat de ser originals en el desenvolupament i/o aplicació d'idees, sovint en un context de recerca.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

CB7. Que els estudiants siguin capaços d'integrar coneixements i enfrontar-se a la complexitat de formular judicis a partir d'una informació que, essent incompleta o limitada, inclogui reflexions sobre les responsabilitats socials i ètiques vinculades a l'aplicació dels seus coneixements i judicis.

CB8. Que els estudiants sàpiguen comunicar les seves conclusions i els coneixements i raons darreres que les sustenten- a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autòdrida o autònoma.

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 algunes eines representatives per a l'aplicació de conceptes teòrics (per exemple, Indyco Builder, PotgreSQL, Oracle, Pentaho Data Integration, Tableau). El curs inclou pràctiques contínues a través d'un projecte de curs, dividit en tres blocs lògics: modelatge de magatzem de dades, integració i migració de dades (ETL), i visualització descriptiva, en què els estudiants treballaran en equips. Hi haurà tres entregables del projecte fora de l'horari de classe, però els estudiants també seran avaluats individualment a l'aula sobre els coneixements adquirits durant cada bloc de el projecte.

OBJECTIUS D'APRENTATGE DE L'ASSIGNATURA

- 1.Ser capaç de modelar magatzems de dades multidimensionals i analitzar les seves dades de forma visual
- 2.Ser capaç d'aplicar tècniques específiques de diseny físic per a sistemes decisionals
- 3.Ser capaç de disenyar i implementar processos de migració de dades (ETL)

HORES TOTALES DE DEDICACIÓ DE L'ESTUDIANTAT

Tipus	Hores	Percentatge
Hores aprenentatge autònom	96,0	64.00
Hores grup petit	27,0	18.00
Hores grup gran	27,0	18.00

Dedicació total: 150 h

CONTINGUTS

Introducció

Descripció:

Comparació entre sistemes operacionals i decisionals; Metadades.

Arquitectures d'emmagatzemament de dades

Descripció:

Corporate Information Factory; DW 2.0



Modelització multidimensional, eines OLAP

Descripció:

Estructura; Restriccions d'Integritat; Operacions; Conceptes avançats

Disseny físic de bases de dades per a consultes analítiques

Descripció:

Star-join i join indexes; Bitmaps; Vistes materialitzades; Dades espai-temporals

Extracció, Transformació i Càrrega

Descripció:

Qualitat de les dades; Integració de dades i esquemes; Gestió de ETL

Visualització i anàlisi descriptiu de la informació

Descripció:

Indicadors de Rendiment Empresarial; Quadres de Comandament

ACTIVITATS

Classes de teoria

Descripció:

En aquestes activitats, el professor introduirà els conceptes teòrics bàsics d'aquesta assignatura. A més de les exposicions, es faran servir tècniques d'aprenentatge cooperatiu. Això requerirà la participació activa dels estudiants i, conseqüentment, seran avaluats.

Objectius específics:

1, 2, 3

Competències relacionades:

CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

CE2. Aplicar els fonaments de la gestió i processament de dades en un problema de ciència de dades

CE5. Modelar, dissenyar i implementar sistemes complexos de dades, incloent-hi la visualització de dades

CE3. Aplicar mètodes d'integració de dades per donar solució a problemes de ciència de dades en entorns heterogenis

CE7. Identificar les limitacions imposades per la qualitat de dades en un problema de ciència de dades i aplicar tècniques per a disminuir el seu impacte

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CT1. EMPRENEDORIA I INNOVACIÓ: Conèixer i comprendre l'organització d'una empresa i les ciències que regeixen la seva activitat; tenir capacitat per entendre les normes laborals i les relacions entre la planificació, les estratègies industrials i comercials, la qualitat i el benefici. Conèixer i entendre els mecanismes en què es basa la recerca científica, així com els mecanismes i instruments de transferència de resultats entre els diferents agents socioeconòmics implicats en els processos d'I+D+i.

CT3. TREBALL EN EQUIP: Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o d'una tasca de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, tot assumint compromisos considerant els recursos disponibles.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autodirigida o autònoma.

CB10. Posseir i comprendre coneixements que aportin una base o oportunitat de ser originals en el desenvolupament i/o aplicació d'idees, sovint en un context de recerca.

CB7. Que els estudiants siguin capaços d'integrar coneixements i enfrontar-se a la complexitat de formular judicis a partir d'una informació que, essent incompleta o limitada, inclogui reflexions sobre les responsabilitats socials i ètiques vinculades a l'aplicació dels seus coneixements i judicis.

CB8. Que els estudiants sàpiguen comunicar les seves conclusions i els coneixements i raons darreres que les sustenten - a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits i la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

Dedicació: 50h

Grup gran/Teoria: 25h

Aprenentatge autònom: 25h



Classes de laboratori

Descripció:

A l'estudiant se li demanarà que practiqui els diferents conceptes introduïts en les classes teòriques. Això inclou resoldre problemes a l'ordinador o en paper.

Objectius específics:

1, 2, 3

Competències relacionades:

CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

CE2. Aplicar els fonaments de la gestió i processament de dades en un problema de ciència de dades

CE5. Modelar, dissenyar i implementar sistemes complexos de dades, incloent-hi la visualització de dades

CE3. Aplicar mètodes d'integració de dades per donar solució a problemes de ciència de dades en entorns heterogenis

CE7. Identificar les limitacions imposades per la qualitat de dades en un problema de ciència de dades i aplicar tècniques per a disminuir el seu impacte

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CT1. EMPRENEDORIA I INNOVACIÓ: Conèixer i comprendre l'organització d'una empresa i les ciències que regeixen la seva activitat; tenir capacitat per entendre les normes laborals i les relacions entre la planificació, les estratègies industrials i comercials, la qualitat i el benefici. Conèixer i entendre els mecanismes en què es basa la recerca científica, així com els mecanismes i instruments de transferència de resultats entre els diferents agents socioeconòmics implicats en els processos d'I+D+i.

CT3. TREBALL EN EQUIP: Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o duent a terme tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, tot assumint compromisos considerant els recursos disponibles.

CB9. Que els estudiants posseïxin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autòdrida o autònoma.

CB10. Posseir i comprendre coneixements que aportin una base o oportunitat de ser originals en el desenvolupament i/o aplicació d'idees, sovint en un context de recerca.

CB7. Que els estudiants siguin capaços d'integrar coneixements i enfrontar-se a la complexitat de formular judicis a partir d'una informació que, essent incompleta o limitada, inclogui reflexions sobre les responsabilitats socials i ètiques vinculades a l'aplicació dels seus coneixements i judicis.

CB8. Que els estudiants sàpiguen comunicar les seves conclusions i els coneixements i raons darreres que les sustenten- a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

Dedicació: 81h

Grup petit/Laboratori: 27h

Aprenentatge autònom: 54h



Examen

Descripció:

Examen escrit dels conceptes teòrics introduïts durant el curs.

Objectius específics:

1, 2, 3

Competències relacionades:

CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

CE2. Aplicar els fonaments de la gestió i processament de dades en un problema de ciència de dades

CE5. Modelar, dissenyar i implementar sistemes complexos de dades, incloent-hi la visualització de dades

CE3. Aplicar mètodes d'integració de dades per donar solució a problemes de ciència de dades en entorns heterogenis

CE7. Identificar les limitacions imposades per la qualitat de dades en un problema de ciència de dades i aplicar tècniques per a disminuir el seu impacte

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CT1. EMPRENEDORIA I INNOVACIÓ: Conèixer i comprendre l'organització d'una empresa i les ciències que regeixen la seva activitat; tenir capacitat per entendre les normes laborals i les relacions entre la planificació, les estratègies industrials i comercials, la qualitat i el benefici. Conèixer i entendre els mecanismes en què es basa la recerca científica, així com els mecanismes i instruments de transferència de resultats entre els diferents agents socioeconòmics implicats en els processos d'I+D+i.

CT3. TREBALL EN EQUIP: Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o duent a terme tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, tot assumint compromisos considerant els recursos disponibles.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autodirigida o autònoma.

CB10. Posseir i comprendre coneixements que aportin una base o oportunitat de ser originals en el desenvolupament i/o aplicació d'idees, sovint en un context de recerca.

CB7. Que els estudiants siguin capaços d'integrar coneixements i enfrontar-se a la complexitat de formular judicis a partir d'una informació que, essent incompleta o limitada, inclogui reflexions sobre les responsabilitats socials i ètiques vinculades a l'aplicació dels seus coneixements i judicis.

CB8. Que els estudiants sàpiguen comunicar les seves conclusions i els coneixements i raons darreres que les sustenten- a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

Dedicació: 19h

Grup gran/Teoria: 2h

Aprenentatge autònom: 17h

SISTEMA DE QUALIFICACIÓ

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

EP = partial (mid term) exam mark

EF = final exam mark

P = Weighted average of the marks of the project deliverables

C = participation in the class



BIBLIOGRAFIA

Bàsica:

- Golfarelli, M.; Rizzi, S. Data warehouse design: modern principles and methodologies. New York [etc.]: McGraw Hill, 2009. ISBN 9780071610391.
- Jensen, C.S.; Pedersen, T.B.; Thomsen, C.W. Multidimensional databases and data warehousing. San Rafael, Calif: Morgan & Claypool, 2010. ISBN 9781608455379.
- Inmon, William H; Imhoff, Claudia; Sousa, Ryan. Corporate information factory. 2nd ed. New York: John Wiley, 2001. ISBN 0471399612.
- Kimball, R. [et al.]. The data warehouse lifecycle toolkit. 2nd ed. Indianapolis: Wiley publishing, 2008. ISBN 9780470149775.
- Garcia-Molina, H.; Ullman, J.D.; Widom, J. Database systems: the complete book. 2nd ed. Harlow, Essex: Pearson Education Limited, 2013. ISBN 9781292037301.
- Vaisman, A.; Zimanyi, E. Data warehouse systems: design and implementation. Berlin: Springer, 2014. ISBN 9783642546549.
- Taniar, D.; Rahayu, W. Data warehousing and analytics: fueling the data engine. Cham: Springer, 2021. ISBN 9783030819781.

Complementària:

- Teorey, Toby ... [et al.]. Database modeling and design: logical design. 5th ed. Burlington, MA: Morgan Kaufmann Publishers/Elsevier, 2011. ISBN 9780123820204.
- Lightstone, S.; Teorey, T.J.; Nadeau, T. Physical database design: the database professional's guide to exploiting indexes, views, storage, and more. Amsterdam [etc.]: Morgan Kaufmann Publishers, 2007. ISBN 9780123693891.
- Ramakrishnan, R.; Gehrke, J. Database management systems. 3rd ed. Boston [etc.]: McGraw-Hill, 2003. ISBN 0071151109.
- Lewis, J. Cost-based oracle fundamentals. Berkeley, CA: Apress, 2006. ISBN 9781590596364.

RECURSOS

Enllaç web:

- <http://cs.ulb.ac.be/conferences/ebiss.html>. Summer school
- <http://tdwi.org>. Transforming Data With Intelligence (former Data Warehouse Institute)
- <https://deds.ulb.ac.be>. MSCA-ITN-Erasmus Joint Doctorate on Data Engineering for Data Science