

**Sistemes Distribuïts**

Codi: 44212  
Crèdits: 6

Titulació	Tipus	Curs	Semestre
4313136 Modelització per a la Ciència i l'Enginyeria / Modelling for Science and Engineering	OT	0	1

**Professor/a de contacte**

Nom: Antonio Espinosa Morales

Correu electrònic: AntonioMiguel.Espinosa@uab.cat

**Equip docent**

Daniel Franco Puntès

**Utilització d'idiomes a l'assignatura**

Llengua vehicular majoritària: anglès (eng)

**Prerequisits**

Es recomana tenir coneixements de programació en Python i conèixer l'ús de sistemes Linux per projectes de desenvolupament.

**Objectius**

Els objectius del mòdul són els següents:

- Donar solucions a problemes d'anàlisi amb eines de codi obert com ara Linux, mysql i Spark
- Entendre les limitacions de les eines de gestió de dades per seleccionar les eines necessàries per a un determinat problema
- Aprendre metodologies de consulta a gestors de dades de cada tecnologia
- Utilitzar eines de Cloud Computing per solucionar problemes d'anàlisi de dades
- Aplicar una metodologia d'anàlisi de dades per resoldre problemes pràctiques

Al final de les sessions de teoria i de laboratori els estudiants haurien de tenir suficients coneixements per entendre els requeriments de un problema d'anàlisi de dades en un contexte industrial. Haurien de poder triar una combinació d'eines i dissenyar una solució per a un problema de dades concret. L'assignatura està orientada a adquirir coneixements de resolució de problemes.

**Competències**

- Analitzar i avaluar arquitectures de computadors paral·lels i distribuïts, així com desenvolupar i optimitzar programari avançat per a les mateixes
- Participar en projectes de recerca i equips de treball en l'àmbit de l'enginyeria de la informació i el còmput d'altres prestacions.
- Que els estudiants sàpiguen comunicar les conclusions pròpies, i els coneixements i les raons últims que les sustenten, a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats.
- Que els estudiants sàpiguen comunicar les seves conclusions, així com els coneixements i les raons últimes que les fonamenten, a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats

- 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.
- Que els estudiants tinguin les habilitats d'aprenentatge que els permetin continuar estudiant, en gran manera, amb treball autònom a autodirigit
- Tenir coneixements que aportin la base o l'oportunitat de ser originals en el desenvolupament o l'aplicació d'idees, sovint en un context de recerca

## Resultats d'aprenentatge

1. Aplicar diverses tècniques de tractament i anàlisi de les dades per preparar aquestes anàlisis en sistemes distribuïts.
2. Aplicar els coneixements adquirits en el disseny de sistemes d'emmagatzematge distribuït, per dissenyar aplicacions intensives de dades i còmput.
3. Conèixer les característiques tècniques de distribució i gestió de dades i les seves implicacions de cost en entorns distribuïts.
4. Que els estudiants sàpiguen comunicar les conclusions pròpies, i els coneixements i les raons últims que les sustenten, a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats.
5. Que els estudiants sàpiguen comunicar les seves conclusions, així com els coneixements i les raons últimes que les fonamenten, a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats
6. 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.
7. Que els estudiants tinguin les habilitats d'aprenentatge que els permetin continuar estudiant, en gran manera, amb treball autònom a autodirigit
8. Seleccionar tant la plataforma distribuïda, com el llenguatge més adequat, a l'hora de generar la proposta de solució a un problema de còmput distribuït.
9. Tenir coneixements que aportin la base o l'oportunitat de ser originals en el desenvolupament o l'aplicació d'idees, sovint en un context de recerca

## Continguts

T1: Introducció als Sistemes Distribuïts y els sistemes de processament de grans volums de dades (4 hores)

T2: Processament de dades en Linux y gestió de workflows (10 hores)

- Arquitectura de sistemes
- Sistemes d'Arxius
- Eines de processament de text
- Gestió de workflows en Linux

T3: Bases de dades relacionals amb Mysql (10 horas)

- Model de dades relacional
- Modelització de dades
- SQL i problem solving amb queries

T4: Processament de dades amb enes Apache(10 horas)

- Limitacions del model relacional amb grans conjunts de dades
- Consistència feble y models relacionats
- Ecosistema d'eines Apache
- Resolver problemes de gestió de dades amb Hadoop, Hive i Spark

T5: Cloud computing (4 hours)

- Introduccio al cloud computing

- Anàlisi de dades en proveedores de Cloud: AWS / Azure

## Metodologia

La metodologia de treball combina treball a classe i sessions de resolució de problemes en les sessions de laboratori

## Activitats formatives

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Laboratori	24	0,96	1, 2, 5, 6, 7, 9
Teoria	38	1,52	1, 2, 3, 4, 6, 8, 9
Tipus: Autònomes			
Desenvolupament d'exercicis pràctics	62	2,48	1, 2, 6, 7

## Avaluació

L'avaluació de l'assignatura es realitzarà amb la combinació del treball desenvolupat en les sessions de laboratori i el examen final

## Activitats d'avaluació

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Examen	30%	2	0,08	1, 2, 5, 9
Laboratori Cloud Computing	10%	4	0,16	1, 3, 6
Laboratori Linux	20%	6	0,24	1, 3, 6, 7
Laboratori Mysql	20%	6	0,24	1, 2, 3, 4, 6, 8, 9
Laboratori Spark	20%	8	0,32	1, 2, 3, 6, 8, 9

## Bibliografia

A. Wittig, M. Wittig. "Amazon Web Services in Action", Manning, 2nd Edition, 2018.

G. Coulouris, J. Dollimore and T. Kinderg, "Distributed Systems. Concepts and design ", Addison-Wesley, 5th edition, 2012.

Bell, Charles; Kindahl, Mats; Thalmann, Lars. "MySQL High Availability". O'Reilly, 2010.

Chang, Fay, et al. "Bigtable: A Distributed Storage System for Structured Data." OSDI, 2006

Dewitt, David, and Jim Gray. "Parallel Database Systems: The Future of High Performance Database Processing." Communications of the ACM 35, no. 6 (1992): 85-98

Schwartz, Baron; Zaitsev, Peter; Tkachenko, Vadim; Zawodny, Jeremy D.; Lentz, Arjen; Balling, Derek J. "High Performance MySQL", O'Reilly, 2008.

Seyed M. M. "Saied" Tahaghoghi and Hugh E. Williams. Learning MySQL. O'Reilly, 2006

Nathan Haines. "Beginning Ubuntu for Windows and Mac Users". Apress 2015. *recurs electrònic a la biblioteca de la UAB*

William E. Shotts. "The Linux Command Line". Second Internet Edition. 2013. <http://linuxcommand.org/tlcl.php>

Petar Zecevic, Marko Bonaci. "Spark in Action". First Edition. Manning. 2017

V. Layka, D. Pollak. "Beginning Scala". Apress. 2015. *recurs electrònic a la biblioteca de la UAB*

Dan C. Marinescu. "Cloud Computing. Theory and Practice". Morgan-Kaufmann. 2018.

R. Buyya, R. N. Calheiros, A. V. Dastjerdi. "Big data. Principles and paradigms". Morgan-Kaufmann. 2016.