

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

Miquel Àngel Senar Rosell

Utilització d'idiomes a l'assignatura

Llengua vehicular majoritària: anglès (eng)

Equip docent extern a la UAB

Pere Pons

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 sàpiguin 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 sàpiguin comunicar les seves conclusions, així com els coneixements i les raons últims que les fonamenten, a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats
- Que els estudiants sàpiguin 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 sàpiguin 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 sàpiguin comunicar les seves conclusions, així com els coneixements i les raons últims que les fonamenten, a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats
6. Que els estudiants sàpiguin 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: Introduction to Distributed Systems and large data processing systems (2 hours)

T2: Cloud computing (2 hours)

- Introduction to cloud computing
- Data analysis with a cloud computing provider: AWS / Azure

T3: Cluster and supercomputer infrastructures (12 hours)

- Principles of job execution under batch queue systems (SLURM).
- Advanced control of jobs: array jobs, dependencies, process binding, heterogeneous resources (GPUs)
- Creation of large jobs and workflows
- Performance engineering principles: tools and methods

T4: Cloud Networking and Virtual Private Clouds (9 hours)

- Intro to VPC
- Build our VPC and launch a web server tutorial
- VPC Lab

T5: Fault tolerance systems (9 hours)

- Availability zones
- Load balancing and autoscaling tutorial
- ELB Lab

T6: Database Cloud project: relational and DynamoDB implementations (9 hours)

- Intro to RDS and DynamoDB
- Build a database server tutorial
- Distributed database Lab

T7: Serverless services and Lambda (9 hours)

- Intro to Lambda services and serverless computing
- Lambda tutorial
- Lambda Lab

Metodologia

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

La metodologia planificada i la proposta d'activitats d'avaluació pot ser modificada depenent de les restriccions a les activitats presencials degut a una actualització de les normatives de salut.

Les classes i els laboratoris virtuals tindran lloc a espais Teams on els estudiants estaran convidats a accedir des del campus virtual. Les sessions de laboratori i la seva planificació serà publicada al principi de curs i s'utilitzaran aquests espais Teams per al desenvolupament de tots els laboratoris pràctics. Els estudiants utilitzaran un entorn computacional propi amb Linux

Nota: es reservaran 15 minuts d'una classe, dins del calendari establert pel centre/titulació, per a la complementació per part de l'alumnat de les enquestes d'avaluació de l'actuació del professorat i d'avaluació de l'assignatura/mòdul.

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
ELB Lab	20%	6	0,24	1, 3, 4, 5, 6, 7
Laboratori Infraestructura	20%	6	0,24	1, 3, 4, 5, 6
Lambda Lab	20%	4	0,16	1, 2, 3, 5, 6, 8, 9
RDS Lab	20%	6	0,24	1, 2, 3, 4, 6, 8, 9
VPC Lab	20%	4	0,16	1, 2, 4, 6, 7

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>

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.

Programari

En l'assignatura es treballarà amb les versions més actualitzades dels sistemes i eines:

-Ubuntu Linux

-SLURM

-Linux development environment