

Titulación	Tipo	Curso
Modelling for Science and Engineering	OP	1

## Contacto

Nombre: Alvaro Wong Gonzalez

Correo electrónico: alvaro.wong@uab.cat

## Equipo docente

Daniel Franco Puentes

## Idiomas de los grupos

Puede consultar esta información al [final](#) del documento.

## Prerrequisitos

Se recomienda tener conocimientos de programación Python y conocer el uso de sistemas Linux para el desarrollo de proyectos.

## Objetivos y contextualización

Los objetivos del módulo son los siguientes:

- Dar soluciones a problemas de análisis de datos con herramientas de código abierto
- Dar soluciones a problemas de análisis de datos como Linux i herramientas de middleware propias de este sistema operativo
- Entender las limitaciones de las herramientas de gestión de datos para seleccionar las herramientas necesarias para un determinado problema
- Aprender metodologías de consulta en gestores de datos de cada tecnología
- Utilizar herramientas de Computación Cloud para solucionar problemas de análisis de datos
- Aplicar una metodología de análisis de datos para resolver problemas prácticos

Al final de las sesiones de teoría y de laboratorio, los estudiantes deberían tener suficientes conocimientos para entender los requerimientos de un problema de análisis de datos en un contexto industrial o académico. Deben poder elegir una combinación de herramientas y diseñar una solución para un problema de datos concreto

## Resultados de aprendizaje

1. CA24 (Competencia) Aplicar las herramientas computacionales del análisis de las grandes bases de datos a la resolución de problemas en el ámbito industrial o de la investigación
2. CA25 (Competencia) Integrar las herramientas computacionales del análisis de las grandes bases de datos en entornos de trabajo multidisciplinares
3. CA26 (Competencia) Trabajar en equipos multidisciplinares con el objetivo de desarrollar proyectos donde se apliquen las técnicas de análisis de grandes bases de datos
4. KA19 (Conocimiento) Caracterizar la estructura y el rendimiento de las diferentes arquitecturas y estructuras de organización y/o gestión de bases de datos
5. KA20 (Conocimiento) Describir las herramientas de computación en la nube utilizadas para el análisis de datos, evaluando sus ventajas y limitaciones.
6. SA24 (Habilidad) Usar software específico en la resolución de problemas de tratamiento de datos en sistemas distribuidos
7. SA25 (Habilidad) Desarrollar aplicaciones informáticas encaminadas a modelizar un proceso concreto mediante sistemas distribuidos, evaluando asimismo el rendimiento computacional de éstas
8. SA26 (Habilidad) Interpretar los informes y resultados que arroja el análisis de una base de datos concreta

## Contenido

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 (14 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
- Virtualization and environments

T4: Cloud Networking and Virtual Private Clouds (8 hours)

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

T5: Fault tolerance systems (4 hours)

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

T6: Database Cloud project: relational and DynamoDB implementations (8 hours)

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

T7: Serverless services and Lambda (2 hours)

- Intro to Lambda services and serverless computing
- Lambda tutorial

- Lambda Lab

## Actividades formativas y Metodología

Título	Horas	ECTS	Resultados de aprendizaje
Tipo: Dirigidas			
Laboratorio	24	0,96	CA24, CA25, SA24, CA24
Teoría	38	1,52	KA19, KA20, SA25, KA19
Tipo: Autónomas			
Desarrollo de ejercicios prácticos	62	2,48	CA26, SA26, CA26

El desarrollo de la asignatura se basará fundamentalmente en el concepto de "aprender haciendo". Habrá sesiones teóricas iniciales para cada tema, en las que el profesor presentará los conceptos claves y se proporcionará a los estudiantes materiales de estudio complementarios (libros, recursos didácticos on-line, artículos y otra documentación técnica). Con esta información, los estudiantes trabajarán en las sesiones prácticas en la resolución de ejercicios y pequeños proyectos de forma individual o en grupos de dos personas. Los alumnos elaborarán informes escritos sobre el trabajo práctico realizado en cada tema.

Nota: se reservarán 15 minutos de una clase dentro del calendario establecido por el centro o por la titulación para que el alumnado rellene las encuestas de evaluación de la actuación del profesorado y de evaluación de la asignatura o módulo.

## Evaluación

### Actividades de evaluación continuada

Título	Peso	Horas	ECTS	Resultados de aprendizaje
ELB Lab	20%	6	0,24	CA26, SA24, SA25, SA26
Laboratorio Infraestructura	30%	6	0,24	CA24, CA26, SA25, SA26
Lambda Lab	10%	4	0,16	CA26, KA20, SA25, SA26
RDS Lab	20%	6	0,24	CA25, CA26, SA25, SA26
VPC Lab	20%	4	0,16	CA26, KA19, SA25, SA26

La evaluación de la asignatura se realizará considerando la combinación del trabajo desarrollado en las sesiones de laboratorio y la entrega de los correspondientes informes.

## Bibliografía

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. *recurso electrónico en 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.

## Software

En la asignatura se trabajará con las versiones más actualizadas de los siguientes sistemas y herramientas:

-Ubuntu Linux

-SLURM

-Linux development environment

## Grupos e idiomas de la asignatura

La información proporcionada es provisional hasta el 30 de noviembre de 2025. A partir de esta fecha, podrá consultar el idioma de cada grupo a través de este [enlace](#). Para acceder a la información, será necesario introducir el CÓDIGO de la asignatura

Nombre	Grupo	Idioma	Semestre	Turno
(PLABm) Prácticas de laboratorio (máster)	1	Inglés	primer cuatrimestre	tarde
(TEm) Teoría (máster)	1	Inglés	primer cuatrimestre	tarde