

## GRAU EN ENGINYERIA INFORMÀTICA DE GESTIÓ I SISTEMES D'INFORMACIÓ

### 103821 - BIG AND OPEN DATA ENGINEERING

#### Informació general

- Tipus d'assignatura : Optativa
- Coordinador : Alfons Palacios González
- Curs: Quart
- Trimestre: Tercer
- Crèdits: 6
- Professorat:
  - Alfons Palacios González <[palacios@tecnocampus.cat](mailto:palacios@tecnocampus.cat)>
  - Laura Rivera Sánchez <[lriviera@tecnocampus.cat](mailto:lriviera@tecnocampus.cat)>

#### Idiomes d'impartició

- Castellà

Aquesta assignatura s'imparteix en castellà.

Tanmateix, la majoria de la documentació amb la que es treballa és en anglès i pot haver algun document també en català.

#### Competències que es treballen

##### Bàsica

- B2\_ Que els estudiants sàpiguen aplicar els seus coneixements a la seva feina o vocació d'una forma professional i tinguin les competències que demostrin mitjançant l'elaboració i defensa d'arguments i la resolució de problemes dins de la seva àrea d'estudi
- B3\_ Que els estudiants tinguin la capacitat de reunir i interpretar dades rellevants (normalment dins de la seva àrea d'estudi), per emetre judicis que incloguin una reflexió sobre temes rellevants de caire social, científica o ètica
- B5\_ Que els estudiants hagin desenvolupat aquelles habilitats d'aprenentatge necessàries per emprendre estudis posteriors amb un alt grau d'autonomia

##### Comú

- CIN1\_Capacitat per a dissenyar, desenvolupar, seleccionar i avaluar aplicacions i sistemes informàtics, assegurant la seva fiabilitat, seguretat i qualitat, segons els principis ètics i la legislació i normativa vigent
- CIN2\_Capacitat per a planificar, concebre, desplegar i dirigir projectes, serveis i sistemes informàtics en tots els àmbits, liderant la seva posada en marxa i la seva millora contínua i valorant el seu impacte econòmic i social
- CIN4\_Capacitat per elaborar el plec de condicions tècniques d'una instal·lació informàtica que compleixi els estàndards i normatives vigents
- CIN5\_Coneixement, administració i manteniment de sistemes, serveis i aplicacions informàtiques
- CIN12\_Coneixement i aplicació de les característiques, funcionalitats i estructura de les bases de dades, que permetin el seu ús adequat i el disseny i l'anàlisi i implementació d'aplicacions basades en elles
- CIN13\_Coneixement i aplicació de les eines necessàries per a l'emmagatzematge, processament i accés als sistemes d'informació, inclosos els basats en web
- CIN14\_Coneixement i aplicació dels principis fonamentals i tècniques bàsiques de la programació paral·lela, concurrent, distribuïda i de temps real

## Específica

- EIS4\_Capacitat d'identificar i analitzar problemes i dissenyar, desenvolupar, implementar, verificar i documentar solucions software sobre la base d'un coneixement adequat de les teories, models i tècniques actuals

## Transversal

- T1\_Que els estudiants coneixin un tercer idioma, que serà preferentment l'anglès, amb un nivell adequat de forma oral i per escrit, d'acord amb les necessitats que tindran les graduades i els graduats a cada titulació

## Descripció

---

Optativa per conèixer les tendències en el tractament de les dades en els camps del Big Data i l'Open Data:

- open data, linked data i controlled vocabularies
- web of data
- big data i big data engineering
- Amb bases de dades relacionals
- Amb bases de dades i datastores NoSQL

## Resultats d'aprenentatge

---

A nivell general, aquesta assignatura contribueix als següents resultats d'aprenentatge especificats per a la matèria d'Enginyeria del software a la qual s'ha d'associar com a optativa:

- (RA1) Utilitzar de forma apropiada teories, procedimientos y herramientas en el desarrollo profesional de la ingeniería informática en todos sus ámbitos (especificación, diseño, implementación, despliegue -implantación- y evaluación de productos) de manera que se demuestre la comprensión de los compromisos adoptados en las decisiones de diseño.
- (RA2) Tomar iniciativas que generen oportunidades, nuevos objetos o soluciones nuevas, con una visión de implementación de proceso y de mercado, y que implique y haga partícipes a los otros en proyectos que se tienen que desarrollar (capacidad de actuar de manera autónoma).
- (RA3) Especificar, diseñar, implementar, gestionar y mantener sistemas y servicios software complejos y/o críticos.
- (RA4) Evaluar y seleccionar plataformas de producción hardware y software para la ejecución de aplicaciones y de servicios informáticos.
- (RA5) Evaluar sistemas hardware/software en función de un criterio de calidad determinado.
- (RA6) Determinar los factores que inciden negativamente en la seguridad y la fiabilidad de un sistema hardware/software, y minimizar sus efectos.
- (RA7) Identificar tecnologías actuales y emergentes y evaluar si son aplicables, y en qué medida, para satisfacer las necesidades de los usuarios.
- (RA8) Diseñar soluciones que integren tecnologías de hardware, software y comunicaciones (y capacidad de desarrollar soluciones específicas de software de sistemas) para sistemas distribuidos y dispositivos de computación ubicua.
- (RA9) Comprender y utilizar eficazmente manuales, especificaciones de productos y otra información de carácter técnico escrita en inglés.

## Metodologia de treball

---

L'enfocament metodològic de l'assignatura és una col·lecció de workshops amb els que introduir-se i provar diferents tecnologies en el tractament de dades massives i/o obertes

Per a cada workshop:

Es fa una introducció el dia interior i es presenta el treball a realitzar al llarg de la setmana abans del dia del workshop

El dia del workshop s'han de realitzar una sèrie de tasques que poden ser d'instal·lació, introducció de dades, programació, recuperació de dades i d'administració dels diferents sistemes

Al començament del següent dia es realitza una prova de validació de l'assoliment de coneixements

A escollir en cada edició de l'assignatura entre cinc i deu workshops entre:

- Big and Data Open Engineering: an Introduction
- Open and Linked Data
- RDF store, Web of Data and controlled vocabularies
- Data Warehousing/BI techniques and ETL Processes
- Big Data Storage with Relational Databases (Oracle 19c o PostgreSQL)
- Big Data with Wide Column Store (Apache Cassandra o DataStax Enterprise)
- Big Data with Key-value Storage (Redis / RedisLabs)
- Big Data with Document Store (MongoDB)
- Big Data Engineering and Architecture with (Apache Hadoop / Apache Spark /Cloudera)
- Big Data with search engines (ElasticSearch)

## Continguts

---

L'assignatura és un tast de tecnologies usades en el camp de les dades massives i/o obertes.

1. Introducció als conceptes Big /Open Data
2. Open Data, Linked Open Data and controlled vocabularies
3. RDF Store, Web of Data and Semantic Web
4. Big data amb sistemes gestors de bases de dades relacionals
5. DataWarehousing, Business Intelligence i processos ETL
6. Big data amb sistemes NoSQL d'emmagatzematge de dades, model key-value (com Redis)
7. Big data amb sistemes NoSQL d'emmagatzematge de dades, model wide-column (com Cassandra)
8. Big data amb sistemes NoSQL d'emmagatzematge de dades, model document-store (com MongoDB)
9. Big data, entorns de desenvolupament per a Data Science i Data Analysis

## Activitats d'aprenentatge

---

Cinc/Deu workshops de quatre/vuit hores entre aquests:

- B&ODE an introduction
- Web of data and semantic web technology
- Linked Open Data and controlled vocabularies
- Big data with RDBMS storage
- DataWarehousing, Business Intelligence and ETL processes
- Big data with wide column storage (Apache Cassandra)
- Big data with key value storage (Redis)
- Big data with document storage (MongoDB)
- Big data with search engines (ElasticSearch)
- Big data, a development environment for data science (Apache Hadoop, Apache Spark, Cloudera)

Cada workshop té una sèrie de tasques a desenvolupar a classe que són avaluades

Cada workshop té una prova d'avaluació

Hi ha un examen final per aquells estudiants que no hagin superat el curs a partir de les qualificacions dels workshops

Cada workshop col·labora als resultats d'aprenentatge RA1..RA9 en el camp de l'emmagatzematge massiu, obert i entrellaçat de les dades

Cada workshop col·labora en les següents competències:

(B2): aplicant els coneixements assolits al llarg del grau per a poder prendre decisions en el camp de les dades massives, obertes i interoperables  
(B3): fent un judici sobre les implicacions socials i ètiques -tant positives com negatives- de la recol·lecció massiva de dades, de l'obertura i la transparència de les dades públiques i dels avantatges i inconvenients de la interoperabilitat de les dades

(B5): proporcionant un bagatge que permeti profunditzar l'estudiant en una o més de les tecnologies presentades al llarg de la seva carrera professional

(T1): usant documentació anglesa

(CIN1): avaluant sistemes d'emmagatzematge de dades

(CIN2): avaluant l'impacte econòmic, social i ètic dels sistemes oberts i/o massius de dades

(CIN4): coneixent les tecnologies emergents en les enginyeries de dades per a prendre decisions en la construcció de sistemes d'informació

(CIN5): coneixent aspectes de l'administració i el manteniment de grans sistemes d'emmagatzematge de dades

(CIN12): coneixent les darreres tendències i les tecnologies emergents en el camp de les bases de dades

(CIN13): coneixent sistemes de programació, introducció i recuperació de dades en sistemes de dades emergents per a dades massives, obertes i interoperables

(CIN14): provant sistemes de distribució i de replicació de dades

(ES14): disposant de criteri per a poder prendre o comprendre decisions sobre tecnologies d'enginyeria de dades

## Sistema d'avaluació

---

Cada workshop representa el 5% (si són deu) de la nota de l'assignatura, en total un 50%. S'avalua el treball previ i el treball fet a classe.

L'examen final representa l'altre 50% de l'assignatura. Un altra forma de realitzar-lo és superar les deu proves de validació dels deu workshops.

## Recursos

---

### Bàsics

#### Audiovisuals

- Bradshaw, Shannon; Chodorow, Kristina "MongoDB: The Definitive Guide". Third Edition. O'Reilly, 2019. ISBN: 978-1491954461.

#### Bibliografies

- Allemang, Dean; Hendler, James; Gandon, Fabien. "Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL". Third Edition. ACM Books, 2020. ISBN 978-1450376143
- Antoniou, Grigoris; Groth, Paul; van Harmelen, Frank; Hoekstra, Rinke. "A Semantic Web Primer". Third Edition. The MIT Press, 2012. ISBN 9780262018289.
- Brown, Mat "Learning Apache Cassandra" . Packt Publishing, 2015. ISBN: 9781783989201.
-

Da Silva, Maxwell Dayvson; Tavares, Hugo Lopes "Redis Essentials". Packt Publishing, 2015. ISBN: 9781784392451.

#### Enllaços web

- W3C (World Wide Web Consortium): W3C's website on Web of Data. (last visited September'2021) <https://www.w3.org/2013/data/>