

**FITXA IDENTIFICATIVA****Dades de l'Assignatura**

Codi	44658
Nom	Big Data
Cicle	Màster
Crèdits ECTS	6.0
Curs acadèmic	2021 - 2022

Titulació/titulacions

Titulació	Centre	Curs	Període
2221 - M.U. en Ciència de Dades	Escola Tècnica Superior d'Enginyeria	1	Segon quadrimestre

Matèries

Titulació	Matèria	Caràcter
2221 - M.U. en Ciència de Dades	10 - Big data	Obligatòria

Coordinació

Nom	Departament
LAPARRA PEREZ-MUELAS, VALERO	242 - Enginyeria Electrònica
LOZANO IBAÑEZ, MIGUEL	240 - Informàtica
SEBASTIAN AGUILAR, RAFAEL	240 - Informàtica

RESUM

Aquesta assignatura es dividix en tres blocs temàtics. En el primer, es presenten els principals problemes associats al processament de grans volums de dades així com les plataformes computacionals capaces de manejar aquests dades eficientment. El primer bloc es tanca revisant casos reals d'èxit derivats de desenrotllaments empresarials en aquest context.

La segona part se centra en la presentació del model MapReduce per a l'explotació de bases de dades sobre plataformes de Big Data.

El tercer bloc s'orienta a l'anàlisi i desenrotllament d'aplicacions basades en aprenentatge màquina sobre plataformes de Big Data. En aquesta part es revisen els llenguatges oferits per les principals plataformes Big Data per a la implementació d'aplicacions, prestant especial atenció a Python i R.



CONEIXEMENTS PREVIS

Relació amb altres assignatures de la mateixa titulació

No heu especificat les restriccions de matrícula amb altres assignatures del pla d'estudis.

Altres tipus de requisits

Anàlisi de senyals
Anàlisi explorador de dades
Aprenentatge màquina (I)
Aprenentatge màquina (II)
Visualització avançada de dades

COMPETÈNCIES

2221 - M.U. en Ciència de Dades

- Que els estudiants sàpiguen comunicar les conclusions (i els coneixements i les raons últimes que les sustenten) a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats.
- Que els estudiants posseïsquen les habilitats d'aprenentatge que els permeten continuar estudiant d'una forma que haurà de ser en gran manera autodirigida o autònoma.
- 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.
- Ser capaços de valorar la necessitat de completar la seua formació tècnica, científica, en llengües, en informàtica, en literatura, en ètica, social i humana en general, i d'organitzar el seu propi autoaprenentatge amb un alt grau d'autonomia
- Capacitat d'accés i gestió de la informació en diferents formats per al seu posterior anàlisi a fi d'obtenir coneixement a partir de dades.
- Ser capaços d'accedir a ferramentes d'informació (bibliogràfiques i d'ocupació) i utilitzar-les apropiadament.
- Ser capaços d'assumir la responsabilitat del seu propi desenvolupament professional i de la seua especialització en un o més camps d'estudi, aplicant els coneixements adquirits en la identificació d'eixides professionals i jaciments d'ocupació.
- Conèixer i utilitzar els distints models d'emmagatzemament de dades i els sistemes de gestió de les bases de dades utilitzant llenguatges de programació de definició, consulta i manipulació dels mateixos.
- Seleccionar, atenent a criteris d'eficiència, escalabilitat, tolerància a fallades i adequació a l'entorn de producció el paradigma de dades òptimes en solucions Big Data. Entendre com les tècniques Big Data s'utilitzen per a suportar i realitzar la presa de decisions basades en dades.



RESULTATS DE L'APRENTATGE

Conéixer el context actual i casos d'èxit en Big Data.

Conéixer i implementar les principals fases emprades en la resolució de problemes d'anàlisi de grans volums de dades.

Conéixer i manejar les principals plataformes i ferramentes per a processar grans volums d'informació.

Processar i analitzar dades numèriques (senyals) i alfanumèrics (text) amb plataformes de Big Data.

Conéixer i implementar el model de MapReduce per a anàlisi de grans volums de dades (bases de dades).

Aplicar els models d'aprenentatge estudiats a problemes de BigData.

DESCRIPCIÓ DE CONTINGUTS

1. Introducció al Big Data

Context, problemes i solucions de BigData

2. Plataformes de BigData

Revisió de les plataformes i ferramentes actuals per a l'anàlisi de grans volums de dades

3. Casos reals: Big Data en l'empresa

Usos de Big Data actuals i en empreses.

4. Plataformes BigData: Hadoop. Mapreduce

1. Entorn de programació i exemples bàsics d'aplicació de MapReduce: desenvolupament i posada en pràctica d'aplicacions.

2. Arquitectura, components i entorn software de Hadoop.

3. Hadoop amb Python.

5. Projectes relacionats amb Hadoop. Kafka

1. Kafka: Sistemes basats en missatges

2. Arquitectura de Kafka (Streaming platform).

3. Desenvolupament d'aplicacions



6. Spark

1. DataFrames, Datasets i SQL.
2. APIs en Spark: PySpark i SparkR.
3. Spark Streaming.
4. Machine Learning Library: Mlib

7. Processament amb big data

Procesament de grans volums de dades al núvol, models de programació i aplicacions orientats a big data (mètodes d'optimització, paral·lelització, gestió de dades, aprenentatge màquina automàtic), arquitectures i sistemes de computació.

8. Exemples

Google Earth Engine, IBM Watson

VOLUM DE TREBALL

ACTIVITAT	Hores	% Presencial
Classes teòricopràctiques	60,00	100
Elaboració de treballs individuals	20,00	0
Estudi i treball autònom	12,00	0
Lectures de material complementari	3,00	0
Preparació d'activitats d'avaluació	12,00	0
Preparació de classes de teoria	20,00	0
Preparació de classes pràctiques i de problemes	13,00	0
Resolució de casos pràctics	10,00	0
TOTAL	150,00	

METODOLOGIA DOCENT

Les classes combinaran la part teòrica amb la pràctica, sense distinció de les sessions dedicades a teoria i pràctica. Totes les sessions s'impartiran en aula d'informàtica.

En la part teòrica s'introduiran els conceptes i mètodes de l'assignatura, acompanyant-los d'exemples i exercicis proposats. Les sessions pràctiques estaran sincronitzades amb la part teòrica i en aquestes l'aprenentatge es realitzarà mitjançant la resolució de problemes, exercicis i casos d'estudi a través dels quals s'adquireixen competències sobre els diferents aspectes de la matèria.



AVALUACIÓ

L'avaluació de l'aprenentatge dels coneixements i competències obtinguts pels estudiants es realitzarà de forma continuada al llarg del curs, i constarà dels següents blocs d'avaluació:

1. Exercicis i treballs entregats durant el curs i/o exàmens parcials: 80% de la nota final.
2. Exàmens: 20% de la nota final.

Les qualificacions obtingudes en l'apartat 1 es conservaran en les dues convocatòries del curs acadèmic en què s'hagen realitzat, donat que la seua avaluació sols és possible en el període de docència.

REFERÈNCIES

Bàsiques

- Bowles, M. (2015). Machine Learning in Python: Essential Techniques for Predictive Analysis. Ed. Wiley
- Wes McKinney, W (2012). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Ipython. Ed. O'Reilly Media
- Scalable Big Data Architecture. Bahaaldine Azarmi. Apress, (2016)
- Amit Nandi. (2015).Spark for Python Developers. Ed. Packt Publishing
- Matei Zaharia, Bill Chambers (2017). Spark: The Definitive Guide. Big data processing made simple. Ed. O'Reilly Media.
- Matei Zaharia, Holden Karau, Andy Konwinski, Patrick Wendell (2015). Learning Spark Lightning-Fast Big Data Analysis. Ed. O'Reilly Media.
- Tom White. (2015) Hadoop: The Definitive Guide. 4th Edition. Storage and Analysis at Internate Scale. O'reilly.
- Zachary Radtko and Donald Miner. (2016) Hadoop with Python. O'reilly
- Kafka: The Definitive Guide: Real-time data and stream processing at scale Neha Narkhede , O'Really 2017

ADDENDA COVID-19

Aquesta addenda només s'activarà si la situació sanitària ho requereix i previ acord del Consell de Govern