

FICHA TÉCNICA DE LA ASIGNATURA

Datos de la asignatura	
Nombre completo	Sistemas de almacenamiento No SQL
Código	DTC-MBD-526
Título	Máster en Big Data. Tecnología y Analítica Avanzada/Master in Big Data Technologies and Advanced Analytics
Impartido en	Máster en Big Data. Tec. y Analítica Avanzada/Master in Big Data Technologies and Advanced Analytics [Primer Curso]
Nivel	Master
Cuatrimestre	Semestral
Créditos	6,0 ECTS
Carácter	Obligatoria
Departamento / Área	Departamento de Telemática y Computación

Datos del profesorado	
Profesor	
Nombre	Diego Yagüe Juárez
Departamento / Área	Departamento de Telemática y Computación
Correo electrónico	dyague@icai.comillas.edu
Profesor	
Nombre	Oswaldo Luis Pérez Ferrer
Departamento / Área	Departamento de Telemática y Computación
Correo electrónico	olperez@icai.comillas.edu

DATOS ESPECÍFICOS DE LA ASIGNATURA

Contextualización de la asignatura
Aportación al perfil profesional de la titulación
<p>The amount of data that is produced and processed worldwide has been growing in the last years as mobile devices and many others such as surveillance cameras, monitor and measuring systems are now part of this continuously evolving world. Also, human interactions, namely social networks, services/product reviews, blogs, etc., have become a very important source of data for the companies or stakeholders to extract value from it.</p> <p>Whit this scenario, the traditional Database systems were found no longer suitable to manage high data volumes (in order of Petabytes) and high data throughput.</p> <p>Currently, BigData use cases are not possible to address without a NoSQL Database. Considering that systems like 5G communications and the increase of electronic or 'smart' appliances, the data volume is expected to grow even more. As a Data Engineer, Data Analyst or Data Scientist, is mandatory to have some skills in NOSQL Databases to leverage their knowledge and design efficient and enriching TI and</p>



Business Solutions.

Prerequisitos

Students willing to take this course should be familiar with OS Virtualization, REPL, Database Query Languages and Java programming principles.

Basic knowledge in Hadoop or similar distributed processing platforms.

Competencias - Objetivos

Competencias

General competences

CG1. Have acquired advanced knowledge and demonstrated, in a research and technological or highly specialized context, a detailed and well-founded understanding of the theoretical and practical aspects, as well as of the work methodology in one or more fields of study.

Haber adquirido conocimientos avanzados y demostrado, en un contexto de investigación científica y tecnológica o altamente especializado, una comprensión detallada y fundamentada de los aspectos teóricos y prácticos y de la metodología de trabajo en uno o más campos de estudio.

CG2. Know how to apply and integrate their knowledge, understanding, scientific rationale, and problem-solving skills to new and imprecisely defined environments, including highly specialized multidisciplinary research and professional contexts.

Saber aplicar e integrar sus conocimientos, la comprensión de estos, su fundamentación científica y sus capacidades de resolución de problemas en entornos nuevos y definidos de forma imprecisa, incluyendo contextos de carácter multidisciplinar tanto investigadores como profesionales altamente especializados.

CG5. Be able to transmit in a clear and unambiguous manner, to specialist and non-specialist audiences, results from scientific and technological research or state-of-the-art innovation, as well as the most relevant foundations that support them.

Saber transmitir de un modo claro y sin ambigüedades, a un público especializado o no, resultados procedentes de la investigación científica y tecnológica o del ámbito de la innovación más avanzada, así como los fundamentos más relevantes sobre los que se sustentan.

CG6. Have developed sufficient autonomy to participate in research projects and scientific or technological collaborations within their thematic area, in interdisciplinary contexts and, where appropriate, with a high knowledge transfer component.

Haber desarrollado la autonomía suficiente para participar en proyectos de investigación y colaboraciones científicas o tecnológicas dentro de su ámbito temático, en contextos interdisciplinares y, en su caso, con una alta componente de transferencia del conocimiento.

CG7. Being able to take responsibility for their own professional development and their specialization in one or more fields of study.

Ser capaces de asumir la responsabilidad de su propio desarrollo profesional y de su especialización en uno o más



campos de estudio.

Specific competences

CE6. Have an overview of the challenges of data management and the importance of choosing the correct database system depending of the use case and technical requirements/limitations.

Tener una visión general de los retos que conlleva el procesamiento y almacenamiento de los datos y la importancia de seleccionar el sistema de base de datos adecuado teniendo en cuenta el caso de uso así como también las características y limitaciones técnicas.

Resultados de Aprendizaje

Learning outcomes

By the end of the course students should:

- RA1. Define the characteristics of a Database System.
- RA2. Understand the advantages of a Relational Data Model and be able to model a real-life problem.
- RA3. Identify the different types of NoSQL databases and their best use cases.
- RA4. Understand the advantages and disadvantages of NoSQL Databases.
- RA5. Understand the problems implicit in distributed processing.
- RA6. Identify the main characteristics and features of the different NoSQL databases.
- RA7. Be able to extract value from the data using the specific features of each database.
- RA8. Be familiar with the different tools to interact, explore or connect to the different NoSQL databases.

BLOQUES TEMÁTICOS Y CONTENIDOS

Contenidos – Bloques Temáticos

Theory

Unit 1. Relational databases

1. What is a database system?
2. Basics of Normalization and Design



3. Applications

Unit 2. Distributed Systems introduction

1. Introduction
2. Characteristics of Distributed Systems
3. Examples and use cases

Unit 3. Non-SQL databases introduction

1. Types of NoSQL DB's
2. CAP Theorem

Unit 4. HBase

1. Architecture
2. Schemas and tables
3. Language Syntax
4. Modeling

Unit 5. MongoDB

1. Architecture
2. JSON Notation
3. Language Syntax
4. Modeling
5. MongoDB Ecosystem
6. Advanced Functions

Unit 6. Neo4J

1. What is a Graph Database?
2. Syntax
3. Modeling

Laboratory

Lab 1. Normalization

In this laboratory students will learn the importance of data consistency, data quality and Data Modeling

Lab 2. HBase

The aim of this laboratory is that students detect data nature and access patterns that will impact directly in the system performance and data modeling

Lab 3. MongoDB

In this session student will learn the flexibility of a schemaless data collection, the importance of data types in a document, basic and advanced syntax, the MongoDB Cloud platform and basic analytics and reporting



Lab 4. Neo4J

In this laboratory student will learn the change of paradigm involved in Graph databases, trying to translate the real-world entities and their relations into an efficient graph data model.

Final project

The final project is nothing more, and nothing less, than an integration activity. Students will put all the modules developed in the previous lab sessions together in order to solve a challenge proposed by the instructor. The topic of the project will be announced at the beginning of the course to allow students to start working on it as soon as possible

METODOLOGÍA DOCENTE

Aspectos metodológicos generales de la asignatura

Inspired by the "learn by doing" paradigm, this course is designed to provide students with the tools they require to get the required skills for modeling, storing and processing data. In every unit, after the initial explanation of each concept, the instructor will propose individual and group quizzes and activities (some of which will be graded) to test students' understanding. Once they are more confident with the material, they will be asked to implement what they have learned in a lab session.

Metodología Presencial: Actividades

In-class activities	Competences
<ul style="list-style-type: none"> Lectures: The lecturer will introduce the fundamental concepts of each unit, along with some practical recommendations, and will go through worked examples to support the explanation. Active participation will be encouraged by raising open questions to foster discussion and by proposing online quizzes and short application exercises to be solved in class either on paper or using a software package. 	CG1, CG7, CE6
<ul style="list-style-type: none"> Lab sessions: Under the instructor's supervision, students individually, will apply the concepts and techniques covered in the lectures to simulated versions of commercial mobile robots. 	CG1, CG2, CG5, CG6, CG7, CE6
<ul style="list-style-type: none"> Tutoring for groups or individual students will be organized upon request. 	–

Metodología No presencial: Actividades

Out-of-class activities	Competences
<ul style="list-style-type: none"> Personal study of the course material and resolution of the proposed exercises. 	CG1, CG7, CE6
<ul style="list-style-type: none"> Lab session preparation to make the most of in-class time. 	CG1



- Lab results analysis and report writing.

CG2, CG5, CE6

RESUMEN HORAS DE TRABAJO DEL ALUMNO

WORK PLAN AND SCHEDULE^[1]

In and out-of-class activities	Date/Periodicity	Deadline
Final exam	After the lecture period	-
Lab sessions	Every 2 sessions	-
Review and self-study of the concepts covered in the lectures	After each lesson	-
Lab preparation	Before every lab session	-
Lab report writing	-	According to the specified date.
Final project	From week 8	Last week

STUDENT WORK-TIME SUMMARY

IN-CLASS HOURS

Lectures	Lab sessions	Assessment
15	13	2

OUT-OF-CLASS HOURS

Self-study	Lab preparation	Lab report writing	Final project
22	4	12	22

ECTS credits:

6 (60 hours)



[1] A detailed work plan of the subject can be found in the course summary sheet (see following page). Nevertheless, this schedule is tentative and may vary to accommodate the rhythm of the class.

EVALUACIÓN Y CRITERIOS DE CALIFICACIÓN

Assessment activities	Grading criteria	Weight
Classroom participation	<ul style="list-style-type: none"> Interest for the contents, challenging questions and proposals. 	10%
Final exam	<ul style="list-style-type: none"> Understanding of the theoretical concepts. Application of these concepts to problem-solving. Critical analysis of numerical exercises' results. 	55%
Lab assignments	<ul style="list-style-type: none"> Application of theoretical concepts to real problem-solving. Ability to model a problem, design data structures and build advanced queries. Written communication skills. 	20%
Final Project	<ul style="list-style-type: none"> Choose the right technology to solve the problem Quality of the proposed solution. Expose the proposed solution using the right technical words 	15%

Calificaciones

Grading

Regular assessment

- Theory** will account for 70%.
- Lab** will account for the remaining 20%.
- Classroom participation** will account 10%

In order to pass the course, the weighted average mark must be greater or equal to 5 out of 10 points, the mark of the final exam must be greater or equal to 5 out of 10 points.

Retake

Lab marks will be preserved as long as the results are a passing grade. Otherwise a new assignment will have to be developed and handed in. In addition, all students will take a final exam. The resulting grade will be computed as follows:



- **Theory** will account for 75.
- **Lab** will account for the remaining 25.

As in the regular assessment period, in order to pass the course, the weighted average mark must be greater or equal to 5 out of 10 points.

Course rules

- Class attendance is mandatory according to Article 93 of the General Regulations (Reglamento General) of Comillas Pontifical University and Article 6 of the Academic Rules (Normas Académicas) of the ICAI School of Engineering. Not complying with this requirement may have the following consequences:
 - Students who fail to attend more than 15% of the lectures may be denied the right to take the final exam during the regular assessment period.
 - Regarding laboratory, absence to more than 15% of the sessions can result in losing the right to take the final exam of the regular assessment period and the retake. Missed sessions must be made up for credit.
- Students who commit an irregularity in any graded activity will receive a mark of zero in the activity and disciplinary procedure will follow (cf. Article 168 of the General Regulations (Reglamento General) of Comillas Pontifical University).

PLAN DE TRABAJO Y CRONOGRAMA

Actividades

Week	In-class activities				Out-of-class activities				Learning outcomes
	Time [h]	Lecture	Laboratory	Assessment	Time [h]	Self-study	Lab preparation and report writing	Other activities	Code
1	2	Course overview (0.5h) Datastores evolution (1.5h)			1	Review and self-study (1h)			RA1, RA2
	2	Introduction to Relational Databases (2h)			1	Review and self-study (1h)			RA1, RA2
		Entity-				Review and			



2	2	Relation Model (1.8)		Quiz (0.2 h)	2	self-study (2h)			RA1, RA2
	2		Lab 1 (2h)		2		Lab preparation (1h) Report writing (3h)		RA1, RA2
3	2	SQL (1.8h)		Quiz (0.2 h)	4	Review and self-study (2h)			RA1, RA2
	2		Lab 2 (2h)		4				RA1, RA2
4	2	Advanced SQL (1.8h)		Quiz (0.2 h)	2	Review and self-study (2h)			RA1, RA2
	2		Lab 3 (2h)		4		Lab preparation (1h) Report writing (3h)		RA1, RA2
5	2	Introduction to NoSQL Databases		Quiz (0.2 h)	4	Review and self-study (2h)			RA3, RA4, RA5
	2		Lab 4 (2h)		4		Lab preparation (1h) Report writing (3h)		RA3, RA4, RA5
	2	Introduction to Hadoop and HBase		Quiz (0.2 h)	4	Review and self-study (2h)			RA3, RA6, RA8



6	2		Lab 5 (2h)		4		Lab preparation (1h) Report writing (3h)		RA3,RA5, RA6
	2	HBase datamodel		Quiz (0.2 h)	6	Review and self-study (2h)			RA3, RA6
7	2		Lab 6 (2h)		6		Lab preparation (1h) Report writing (3h)		RA3, RA5, RA6
	2	HBase Arquitecture 1		Quiz (0.2 h)	4	Review and self-study (2h)			RA3, RA5, RA6, RA7
8	2		Lab 7 (2h)		8		Lab preparation (1h) Report writing (3h)		RA3, RA6, RA7, RA8
	2	HBase Arquitecture 2		Quiz (0.2 h)		Review and self-study (2h)			RA3, RA5, RA6, RA7
9	2		Lab 8 (2h)				Lab preparation (1h) Report writing (3h)		RA3, RA6, RA7, RA8
	2	HBase Design		Quiz (0.2 h)		Review and self-study		Final project development (2h)	RA6, RA7, RA8



10					study (2h)			
	2		Lab 9 (2h)			Lab preparation (1h) Report writing (3h)		RA6, RA7, RA8
11	2	Introduction MongoDB		Quiz (0.2 h)	Review and self-study (2h)			RA3, RA5, RA8
	2		Lab 10 (2h)			Lab preparation (1h) Report writing (3h)		RA3, RA5, RA8
12	2	MongoDB Arquitecture		Quiz (0.2 h)	Review and self-study (2h)		Final project development (2h)	RA3, RA7, RA8
	2		Lab 11 (2h)			Lab preparation (1h) Report writing (3h)		RA3, RA7, RA8
13	2	Introduction to Neo4J		Quiz (0.2 h)	Review and self-study (2h)			RA3, RA5, RA8
	2		Lab 12 (2h)			Lab preparation (1h) Report writing (3h)		RA3, RA5, RA8
					Review			



14	2	Arquitecting Neo4J (1h)	Lab 13 (1h)	Quiz (0.2 h)		and self-study (2h)			RA3, RA7, RA8
	2		Final project (2h)					Final project development (6h)	
15	2		Final project (2h)					Final project development (6h)	
	2		Final project demonstration (2h)					Final project presentation preparation (4h)	

BIBLIOGRAFÍA Y RECURSOS

Bibliografía Básica

- Slides prepared by the lecturer (available in Moodlerooms)
- Shripav , Shashwat - Learning HBase .Birmingham, Packt Publishing.
- George, Lars (2011): HBase : The Definitive Guide. Sebastopol,O'Reilly
- Banker , Kyle, Peter Bakkum , Shaun Verch ,Douglas Garrett, Tim Hawkins (2016): MongoDB in Action . Second Edition . New York, Manning Publications Co.
- Vukotic , Aleksa , Nicki Watt (2015): Neo4j in Action . New York, Manning Publications

Bibliografía Complementaria

- Spaggiari, Jean-Marc, Kevin O'Dell (2016): Architecting HBase Applications . Sebastopol,O'Reilly.
- <http://hbase.apache.org/book.html>
- <https://docs.mongodb.com/>
- Panzarino , Onofrio (2014): Learning Cypher. Birmingham , Packt Publishing
- <https://neo4j.com/docs/javareference/current/>

En cumplimiento de la normativa vigente en materia de **protección de datos de carácter personal**, le informamos y recordamos que puede consultar los aspectos relativos a privacidad y protección de datos que ha aceptado en su matrícula entrando en esta web y pulsando "descargar"

<https://servicios.upcomillas.es/sedelectronica/inicio.aspx?csv=02E4557CAA66F4A81663AD10CED66792>