



UNIVERSIDAD  
POLITÉCNICA  
DE MADRID

PROCESO DE  
COORDINACIÓN DE LAS  
ENSEÑANZAS PR/CL/001



E.T.S. de Ingenieros  
Informaticos

# ANX-PR/CL/001-01

## GUÍA DE APRENDIZAJE

### ASIGNATURA

105000416 - Arquitectura Del Data Warehouse

### PLAN DE ESTUDIOS

10ID - Doble Grado En Ingenieria Informatica Y En Ade

### CURSO ACADÉMICO Y SEMESTRE

2022/23 - Segundo semestre

## Índice

---

### Guía de Aprendizaje

1. Datos descriptivos.....	1
2. Profesorado.....	1
3. Conocimientos previos recomendados.....	2
4. Competencias y resultados de aprendizaje.....	3
5. Descripción de la asignatura y temario.....	5
6. Cronograma.....	7
7. Actividades y criterios de evaluación.....	9
8. Recursos didácticos.....	13
9. Otra información.....	16

## 1. Datos descriptivos

### 1.1. Datos de la asignatura

<b>Nombre de la asignatura</b>	105000416 - Arquitectura del Data Warehouse
<b>No de créditos</b>	3 ECTS
<b>Carácter</b>	Obligatoria
<b>Curso</b>	Cuarto curso
<b>Semestre</b>	Octavo semestre
<b>Período de impartición</b>	Febrero-Junio
<b>Idioma de impartición</b>	Castellano
<b>Titulación</b>	10ID - Doble Grado en Ingeniería Informática y en ADE
<b>Centro responsable de la titulación</b>	10 - Escuela Técnica Superior De Ingenieros Informaticos
<b>Curso académico</b>	2022-23

## 2. Profesorado

### 2.1. Profesorado implicado en la docencia

<b>Nombre</b>	<b>Despacho</b>	<b>Correo electrónico</b>	<b>Horario de tutorías *</b>
Santiago Eibe Garcia (Coordinador/a)	2311	santiago.eibe@upm.es	Sin horario. Consultar las tutorías en el Aula Virtual de la asignatura. Concertar cita previa mediante correo electrónico (seibe@fi.upm.es)

Raul Alonso Calvo	2315	raul.alonso@upm.es	Sin horario. Sin horario. Consultar las tutorías en el Aula Virtual de la asignatura
-------------------	------	--------------------	---

\* Las horas de tutoría son orientativas y pueden sufrir modificaciones. Se deberá confirmar los horarios de tutorías con el profesorado.

### 3. Conocimientos previos recomendados

---

#### 3.1. Asignaturas previas que se recomienda haber cursado

El plan de estudios Doble Grado en Ingeniería Informática y en ADE no tiene definidas asignaturas previas recomendadas para esta asignatura.

#### 3.2. Otros conocimientos previos recomendados para cursar la asignatura

- Sistemas de Virtualización Ligera basados en contenedores (Docker)
- IDE: Eclipse, IntelliJ IDEA o similar
- Conocimientos básicos de administración/configuración de sistemas y redes TCP/IP
- Bases de Datos Relacionales: lenguaje SQL y SGBD MySQL
- Programación en lenguaje Java
- Ofimática básica

## 4. Competencias y resultados de aprendizaje

---

### 4.1. Competencias

30AD-CB03 - Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.

30AD-CE38 - Aplicar distintos tipos de Sistemas de Información que apoyan a la gestión empresarial.

10II-CE05 - Capacidad de diseñar y realizar experimentos apropiados, interpretar los datos y extraer conclusiones.

10II-CE06 - Comprender intelectualmente el papel central que tienen los algoritmos y las estructuras de datos, así como una apreciación del mismo.

10II-CE07 - Entender el soporte físico (hardware) de los ordenadores desde el punto de vista del soporte lógico (software), por ejemplo, el uso del procesador, de la memoria, de los discos, del monitor, etc.

10II-CE13/18 - Comprender lo que pueden y no pueden conseguir las tecnologías actuales, y las limitaciones de la informática, que implica distinguir entre lo que, inherentemente, la informática no es capaz de hacer y lo que puede lograrse a través de la ciencia y la tecnología futuras.

10II-CE14/015 - Conocer el software, el hardware y las aplicaciones existentes en el mercado, así como el uso de sus elementos, y capacidad para familiarizarse con nuevas aplicaciones informáticas.

10II-CE44 - Conocimiento de tecnologías punteras relevantes y su aplicación.

10II-CE48 - Gestionar sistemas y servicios informáticos en contextos empresariales o institucionales para mejorar sus procesos de negocio.

10II-CE53/54 - Capacidad para trabajar de forma efectiva como individuo, organizando y planificando su propio trabajo, de forma independiente o como miembro de un equipo.

10II-CE56 - Ser capaz de aclarar la relevancia y utilidad de la teoría y las habilidades aprendidas en el contexto académico sobre los acontecimientos del mundo real.

10II-CG01/21 - Capacidad de resolución de problemas aplicando conocimientos de matemáticas, ciencias e ingeniería.

10II-CG02/CE45 - Capacidad para el aprendizaje autónomo y la actualización de conocimientos, y reconocimiento

de su necesidad en el área de la informática.

10II-CG03/04 - Saber trabajar en situaciones carentes de información y bajo presión, teniendo nuevas ideas, siendo creativo.

10II-CG05 - Capacidad de gestión de la información.

10II-CG19 - Capacidad para usar las tecnologías de la información y la comunicación.

30AD-CG11 - Que los estudiantes sean capaces de utilizar las nuevas tecnologías de la información y las comunicaciones aplicándolas al ámbito de la dirección y administración de empresas.

## 4.2. Resultados del aprendizaje

RA165 - Ser capaz de recoger datos, ordenarlos e interpretarlos.

RA145 - Buscar información, su análisis, interpretación, síntesis y transmisión.

RA198 - Tener capacidad de razonamiento, abstracción y resolución de problemas.

RA199 - Interpretar las diferentes soluciones y obtener las conclusiones apropiadas a los contextos considerados.

RA197 - Aplicar métodos sistemáticos para tomar decisiones personales con coherencia, acierto y seguridad.

RA195 - Conocer las tecnologías para la ayuda a toma de decisiones empresariales.

RA196 - Relacionar el sistema de información con la estrategia de la empresa.

RA194 - Aplicar diferentes categorías de sistemas de información.

RA146 - Tener capacidad de razonamiento y abstracción.

RA149 - Poder exponer y comunicar las soluciones a los problemas tanto de forma oral como escrita.

## 5. Descripción de la asignatura y temario

---

### 5.1. Descripción de la asignatura

El principal objetivo de la asignatura es presentar la problemática asociada con la implantación de un **Data Warehouse** y/o una infraestructura de **Big Data** tanto desde el punto de vista de la arquitectura propiamente dicha como de las tecnologías sobre las que se sustenta y con las que interacciona. Por ende, se persigue exponer la problemática, más general, de los modernos sistemas de gestión de datos en la actualidad en los que tanto el *Data Warehouse* como el *Big Data* desempeñan un papel protagonista. En este sentido, la segunda parte del curso se focaliza más en la problemática del *Big Data* y del ecosistema de *Hadoop*

### 5.2. Temario de la asignatura

1. Problemática Actual de los Sistemas de Gestión de Datos
  - 1.1. Revisión de la Arquitectura Tradicional de los Sistemas Gestores de Bases de Datos
  - 1.2. Modelo Relacional. Funcionalidades básicas de un SGBD(R)
  - 1.3. Introducción NoSQL
2. Almacenamiento de Datos. Problema de Volumen Masivo de Datos
  - 2.1. Jerarquía de Almacenamiento
  - 2.2. Problemática de Rendimiento. Técnicas de Indexación
3. Procesamiento de Interrogaciones. OLTP vs OLAP
  - 3.1. Preprocesamiento y Análisis Sintáctico. Árboles de Interrogación
  - 3.2. Análisis de costes. Estimación
  - 3.3. Plan de ejecución. Operadores
4. Caso de Estudio: Relacional o No Relacional (esa es la cuestión)
  - 4.1. Arquitectura y Administración Básica
  - 4.2. Concepto de Motor de Almacenamiento
  - 4.3. Funcionalidad PostRelacional
  - 4.4. Comparativa MySQL/MariaDB vs NoSQL
5. Introducción al Data Warehouse
  - 5.1. Fundamentos del Data Warehouse. OLAP

- 5.2. Arquitectura de un Data Warehouse
- 5.3. Módulos y Funcionalidad de un Data Warehouse
- 5.4. Técnicas y Herramientas OLAP
- 5.5. Caso de Estudio: Pentaho BI Server
- 6. ETLs
  - 6.1. Funcionalidad ETL
  - 6.2. Taxonomía: ETL vs ELT vs Streaming ETL
  - 6.3. Metadatos. Integración. Automatización de Procesos
  - 6.4. Estudio de Herramientas ETL. Spoon, Talend
- 7. Base de Datos del Data Warehouse. Modelo Multidimensional
  - 7.1. Diseño Multidimensional Básico
  - 7.2. Diseño Multidimensional Extendido
  - 7.3. MOLAP, ROLAP, HOLAP
- 8. Metodologías de Diseño del Data Warehouse
  - 8.1. Técnicas de Diseño del Data Warehouse
  - 8.2. Bus Común del Data Warehouse
  - 8.3. DW 2.0
- 9. Alta Disponibilidad y Heterogeneidad en el Data Warehouse
  - 9.1. Particionado de Datos
  - 9.2. Replicación de Datos
- 10. Big Data
  - 10.1. Ecosistema de Hadoop. Hive
  - 10.2. Streaming. Kafka

## 6. Cronograma

### 6.1. Cronograma de la asignatura \*

Sem	Actividad en aula	Actividad en laboratorio	Tele-enseñanza	Actividades de evaluación
1	<b>1 Sistemas de Gestión de Datos</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
2	<b>2 Almacenamiento de Volúmenes Masivos de Datos</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
3	<b>3 Procesamiento de Interrogaciones: OLTP vs OLAP</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
4		<b>4 Caso Práctico: MySQL/MariaDB vs NoSQL</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
5		<b>4 Caso Práctico: MySQL/MariaDB Dockerizado</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
6	<b>5.1 Introducción al Data Warehouse. OLAP</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			<b>Tecnologías en el DW</b> TG: Técnica del tipo Trabajo en Grupo Evaluación continua No presencial Duración: 20:00
7	<b>5.2 Arquitectura del Data Warehouse</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
8	<b>5.3 Módulos y Funcionalidad de un Data Warehouse</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
9	<b>6 ETLs</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
10	<b>7 Modelo Multidimensional</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral			<b>Integración de procesos en el DW: ETLs</b> TG: Técnica del tipo Trabajo en Grupo Evaluación continua No presencial Duración: 28:00
	<b>8 Metodologías de Diseño del Data Warehouse</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral			

11	<b>9 Alta Disponibilidad y Heterogeneidad en el Data Warehouse</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
12	<b>10.1 Ecosistema de Hadoop</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
13	<b>10.1 Hive</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
14	<b>10.2 Streaming. Kafka</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
15	<b>10.2 Streaming. Kafka</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			<b>Tecnologías Data Warehouse y Big Data</b> TI: Técnica del tipo Trabajo Individual Evaluación sólo prueba final No presencial Duración: 12:00
16				<b>Nota de Participación NPa</b> OT: Otras técnicas evaluativas Evaluación continua y sólo prueba final No presencial Duración: 00:00
17				<b>Examen Individual (EP)</b> EX: Técnica del tipo Examen Escrito Evaluación continua Presencial Duración: 02:00  <b>Examen Individual (EG)</b> EX: Técnica del tipo Examen Escrito Evaluación sólo prueba final Presencial Duración: 02:00

Para el cálculo de los valores totales, se estima que por cada crédito ECTS el alumno dedicará dependiendo del plan de estudios, entre 26 y 27 horas de trabajo presencial y no presencial.

\* El cronograma sigue una planificación teórica de la asignatura y puede sufrir modificaciones durante el curso derivadas de la situación creada por la COVID-19.

## 7. Actividades y criterios de evaluación

### 7.1. Actividades de evaluación de la asignatura

#### 7.1.1. Evaluación (progresiva)

Sem.	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
6	Tecnologías en el DW	TG: Técnica del tipo Trabajo en Grupo	No Presencial	20:00	25%	3.5 / 10	10II-CE13/18 10II-CE48 30AD-CG11 10II-CE44 10II-CG02/CE45 30AD-CE38
10	Integración de procesos en el DW: ETLs	TG: Técnica del tipo Trabajo en Grupo	No Presencial	28:00	35%	3.5 / 10	10II-CE44 10II-CG05 10II-CE13/18 10II-CE05 10II-CE07 30AD-CB03 10II-CE06
16	Nota de Participación NPa	OT: Otras técnicas evaluativas	No Presencial	00:00	15%	0 / 10	10II-CE44 10II-CG02/CE45 10II-CG05 10II-CE13/18 10II-CE05 10II-CE48 10II-CE14/015 10II-CE56 10II-CE07 10II-CE53/54 10II-CG19 30AD-CG11 30AD-CB03 30AD-CE38 10II-CE06 10II-CG01/21
17	Examen Individual (EP)	EX: Técnica del tipo Examen Escrito	Presencial	02:00	25%	5 / 10	10II-CE56 10II-CE53/54 10II-CG01/21 10II-CG03/04

#### 7.1.2. Prueba evaluación global

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
15	Tecnologías Data Warehouse y Big Data	TI: Técnica del tipo Trabajo Individual	No Presencial	12:00	30%	5 / 10	10II-CE14/015 10II-CG19 30AD-CG11 10II-CG02/CE45 10II-CE13/18
16	Nota de Participación NPa	OT: Otras técnicas evaluativas	No Presencial	00:00	15%	0 / 10	10II-CE44 10II-CG02/CE45 10II-CG05 10II-CE13/18 10II-CE05 10II-CE48 10II-CE14/015 10II-CE56 10II-CE07 10II-CE53/54 10II-CG19 30AD-CG11 30AD-CB03 30AD-CE38 10II-CE06 10II-CG01/21
17	Examen Individual (EG)	EX: Técnica del tipo Examen Escrito	Presencial	02:00	55%	5 / 10	10II-CE56 10II-CE53/54 10II-CG01/21 10II-CG03/04

### 7.1.3. Evaluación convocatoria extraordinaria

Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
Examen convocatoria extraordinaria	EX: Técnica del tipo Examen Escrito	Presencial	02:00	100%	5 / 10	10II-CE56 10II-CG01/21 10II-CG03/04

## 7.2. Criterios de evaluación

Se plantean las siguientes modalidades de evaluación, denominadas progresiva (**EP**) y global (**EG**), excluyentes entre si:

### 1. Evaluación Progresiva (EP)

La asignatura se evaluará preferentemente de forma progresiva mediante **2 proyectos colaborativos** más un **examen individual** de acuerdo con lo siguiente:

1. Los proyectos en grupo son actividades colaborativas que se realizarán obligatoriamente durante el período lectivo y en grupos de hasta 4 alumnos de entre los matriculados de la asignatura
2. Tanto la monitorización como la retroalimentación de los proyectos se harán preferentemente en la plataforma Moodle pero también en horario de clase valorándose específicamente la participación de los alumnos
3. Dicha participación en las actividades de la asignatura, tanto en el aula como a través de la plataforma de teleenseñanza, se valorará como **NPa**. El peso de este ítem es del **15%** del total
4. Los proyectos suponen el **60%** del total de la nota de la **EP** (ver tabla de evaluación sumativa apartado 7.1.1)
5. Es obligatorio realizar todos los proyectos y alcanzar en cada uno de ellos una calificación igual o superior al **35%** de la valoración total del proyecto. En caso de no poder entregar alguno de los proyectos se requerirá documentación que lo justifique
6. Solamente en el caso de que en los proyectos se haya alcanzado una nota igual o superior al requisito mínimo el alumno realizará al final del curso un **examen individual** cuyo objetivo es doble:
  - Por un lado se pretende evaluar si se han alcanzado un nivel mínimo de conocimientos en los contenidos básicos de la asignatura y/o aquéllos no cubiertos suficientemente por los proyectos
  - En segundo lugar se persigue evaluar la aportación individual del alumno a los proyectos realizados conjuntamente con otros compañeros durante el período lectivo

### 2. Evaluación Global (EG)

Si no se ha superado la evaluación progresiva debido a que:

- por razones debidamente justificadas no se hubiera entregado alguno de los proyectos
- porque no se satisface el criterio mínimo en la evaluación de los mismos

el alumno realizará:

- un proyecto individual de recuperación de la nota de proyectos cuyo peso es del **30%** en la evaluación sumativa (ver tabla de evaluación sumativa apartado 7.1.2)
- una prueba global individual cuyo peso es del **55%** en la evaluación sumativa (ver tabla de evaluación sumativa apartado 7.1.2)

## Resumen Ordinaria/Extraordinaria

Para superar la asignatura en la **convocatoria ordinaria de junio** se establecen dos situaciones posibles:

1. Obtener un mínimo de 50 puntos sobre los 100 disponibles en el cómputo global de la EP
2. De forma totalmente **excluyente** con la anterior (sólo si el alumno no ha superado la EP) obtener una nota mínima igual o superior al 50% de la valoración en la EG

Para poder superar la asignatura en la **convocatoria extraordinaria de julio** se establecen los siguientes requisitos:

1. No se evaluará mediante proyectos sino que únicamente se realizará un examen individual que cubrirá todos los aspectos teóricos y prácticos de la asignatura
2. En esta prueba se deberá obtener un mínimo de 50 puntos sobre los 100 disponibles en el cómputo global

## Actuación ante comportamientos fraudulentos

Dada la naturaleza de los conocimientos y tecnologías abarcadas en esta asignatura, se plantea el problema de la existencia de innumerables fuentes de información a disposición, desde ideas a desarrollar, pasando por códigos de todo tipo hasta aplicaciones completas. Es por esto que se premiará especialmente la originalidad y el esfuerzo propios, sobre el uso de materiales no propios. El uso de materiales ajenos de cualquier naturaleza (código, ideas, etc.) deberá ser debidamente declarado públicamente e identificado claramente, reconociendo su extensión y citando las fuentes de su autoría original. En caso contrario se considerará como plagio

Si se detecta plagio en algún proyecto, los alumnos involucrados perderán la nota que hubiera obtenido con anterioridad

## Actuación ante detección de fraudes o copias/plagio

Los derechos y deberes de los estudiantes universitarios están desarrollados en los Estatutos de la Universidad Politécnica de Madrid (BOCM de 15 de noviembre de 2010) y en el Estatuto del Estudiante Universitario (RD

1791/2010 de 30 de diciembre). El artículo 124 a) de los Estatutos de la UPM fija como deber del estudiante "Seguir con responsabilidad y aprovechamiento el proceso de formación, adquisición de conocimientos, y aprendizaje correspondiente a su condición de universitario" y el artículo 13 del Estatuto del Estudiante Universitario, en el punto d) especifica también como deber del estudiante universitario "abstenerse de la utilización o cooperación en procedimientos fraudulentos en las pruebas de evaluación, en los trabajos que se realicen o en documentos oficiales de la universidad"

En el caso de que en el desarrollo de las pruebas de evaluación se aprecie el incumplimiento de los deberes como estudiante universitario, el coordinador de la asignatura podrá ponerlo en conocimiento del Director o Decano del Centro, que de acuerdo con lo establecido en el artículo 74 (n) de los Estatutos de la UPM tiene competencias para "Proponer la iniciación del procedimiento disciplinario a cualquier miembro de la Escuela o Facultad, por propia iniciativa o a instancia de la Comisión de Gobierno" al Rector, en los términos previstos en los estatutos y normas de aplicación

## 8. Recursos didácticos

---

### 8.1. Recursos didácticos de la asignatura

Nombre	Tipo	Observaciones
Building the Data Warehouse. W.H. Immon. 1996. Willey	Bibliografía	
Managing the Data Warehouse. W.H. Immune. 1997. Willey	Bibliografía	
Building the Operational Data Store. W. H. Immon. 1999. Willey	Bibliografía	
Exploration Datawarehouse. W. Immon. 2000. Willey	Bibliografía	
The Data Warehouse lifecycle toolkit. R. Kimball. 2000. Willey	Bibliografía	

Improving Data Warehouse and Business Information Quality. Methods for reducing cost and increasing profits. L. English. 1999 Willey	Bibliografía	
Principles of Data Base Systems (Second Edition), Jeffrey D. Ullman, Ed. Computer Science Press, Rockville, Maryland, 1982	Bibliografía	
First Course in Database Systems, Jeffrey D. Ullman, Jennifer Widom, ISBN-10: 013600637X. 2007	Bibliografía	
Sistemas de Bases de Datos, R. Elmasri y S.B.Navathe, 2ª edición, Addison-Wesley Iberoamericana, 1997	Bibliografía	
Fundamentos de Bases de Datos, A. Silberschatz, H. Korth, S. Sudarsham, 5ª edición, Mcgraw-Hill, 2006	Bibliografía	
Database Systems-A Practical Approach to Design, Implementation and Management. 4th ed., Connolly, T., Begg, C. AND Strachan, A., 2004. Addison-Wesley	Bibliografía	
Fundamentals of Database Systems, 5th ed ., Elmasri, R. Navathe, S., 2006.. Addison-Wesley	Bibliografía	
MySQL Administrator's Bible, Sheeri Cabral, Keith Murphy. Wiley Publishing 2009	Bibliografía	
The Complete Reference MySQL, Vikram Vaswani.McGraw-Hill/Osborne 2007	Bibliografía	

Database Systems: The Complete Book, Hector Garcia-Molina, Jeff Ullman, and Jennifer Widom. (DS-CB), 2008, 2nd edition	Bibliografía	
Mondrian 3.0 Technical Guide	Bibliografía	
Pentaho Data Integration 2ªEd María Carina Roldán	Bibliografía	
Pentaho Data Integration 4 Cookbook A.S.Pulvirenti, M.C. Roldán	Bibliografía	
Hadoop, The Definitive Guide T. White, 4ª edición, O'Reilly, 2015	Bibliografía	
Architecting Data Lakes 2Ed. Ben Sharma. O'Reilly Media 2018	Bibliografía	
Architecting Modern Data Platforms. Kunigk. O'Really 2018	Bibliografía	
Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems. Martin Kleppmann. O'Really 2017	Bibliografía	
The Enterprise Big Data Lake: Delivering the Promise of Big Data and Data Science. Alex Gorelik. O'Really 2019	Bibliografía	

## 9. Otra información

---

### 9.1. Otra información sobre la asignatura