



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

105000167 - Arquitectura del Data Warehouse

PLAN DE ESTUDIOS

10II - Grado En Ingenieria Informatica

CURSO ACADÉMICO Y SEMESTRE

2019/20 - 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.....	4
6. Cronograma.....	6
7. Actividades y criterios de evaluación.....	8
8. Recursos didácticos.....	10

1. Datos descriptivos

1.1. Datos de la asignatura

Nombre de la asignatura	105000167 - Arquitectura del Data Warehouse
No de créditos	3 ECTS
Carácter	Optativa
Curso	Cuarto curso
Semestre	Octavo semestre
Período de impartición	Febrero-Junio
Idioma de impartición	Castellano
Titulación	10II - Grado En Ingenieria Informatica
Centro responsable de la titulación	10 - Escuela Tecnica Superior de Ingenieros Informaticos
Curso académico	2019-20

2. Profesorado

2.1. Profesorado implicado en la docencia

Nombre	Despacho	Correo electrónico	Horario de tutorías *
Santiago Eibe Garcia (Coordinador/a)	2311	santiago.eibe@upm.es	L - 12:00 - 13:30 L - 15:30 - 17:00 X - 12:00 - 13:30 J - 15:30 - 17:00 Concertar cita previa mediante correo electrónico (seibe@fi.upm.es)

Raul Alonso Calvo	2315	raul.alonso@upm.es	L - 10:00 - 13:00 X - 10:00 - 13:00
-------------------	------	--------------------	--

* 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

- Bases De Datos

3.2. Otros conocimientos previos recomendados para cursar la asignatura

- Conocimiento y experiencia con MySQL en particular
- Conocimientos básicos de administración/configuración de sistemas y redes TCP/IP
- En particular, establecimiento y utilización de conexiones mediante túneles ssh
- Programación en lenguaje Java
- Ofimática básica
- Conocimientos de bases de datos
- Conocimientos de SQL

4. Competencias y resultados de aprendizaje

4.1. Competencias

CG-2/CE45 - Capacidad para el aprendizaje autónomo y la actualización de conocimientos, y reconocimiento de su necesidad en el área de la informática.

Ce 12/16 - Conocer los campos de aplicación de la informática, y tener una apreciación de la necesidad de poseer unos conocimientos técnicos profundos en ciertas áreas de aplicación; apreciación del grado de esta necesidad en, por lo menos, una situación.

Ce 13/18 - 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.

Ce 44 - Conocimiento de tecnologías punteras relevantes y su aplicación.

4.2. Resultados del aprendizaje

RA276 - Dado un campo de aplicación de la informática, evaluar y diseñar el sistema informático más apropiado para resolver alguno de sus problemas, exponiendo las dificultades técnicas y los límites de la aplicación.

RA373 - Aplicar las técnicas y métodos relativos a una línea de especialización concreta del área tecnológica, comprendiendo sus límites tanto teóricos como prácticos, para la resolución de un problema o necesidad planteado por un consumidor o cliente real

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 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 sistemas de gestión de datos en la actualidad tomando para ello el DW como hilo conductor

En la parte final se aborda una introducción a Big Data y al ecosistema de Hadoop

5.2. Temario de la asignatura

1. Sistemas gestores de bases de datos
 - 1.1. Introducción a SGBD
 - 1.2. Funcionalidades básicas de un SGBD
2. Almacenamiento de datos
 - 2.1. Jerarquía de almacenamiento
 - 2.2. Técnicas de indexación. Rendimiento
3. Procesamiento de interrogaciones
 - 3.1. Preprocesamiento y análisis sintáctico. Árboles de interrogación
 - 3.2. Análisis de costes. Estimación
 - 3.3. Plan de ejecución. Operadores
4. Caso práctico de MySQL
 - 4.1. Arquitectura y administración básica
 - 4.2. Gestión de almacenamiento
 - 4.3. Control de acceso
5. Introducción al Data Warehouse
 - 5.1. Concepto básicos del Data Warehouse
 - 5.2. Arquitectura de un Data Warehouse
 - 5.3. Módulos y Funcionalidad de un Data Warehouse

6. ETL

6.1. Funcionalidad ETL

6.2. Herramientas ETL

7. Base de Datos del Data Warehouse

7.1. Diseño multidimensional básico

7.2. Diseño multidimensional extendido

7.3. Sistemas gestores de bases de datos para el data warehouse

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. Replicación y particionado

9.1. Alta disponibilidad y Heterogeneidad en el Data Warehouse

10. Consulta de información del Data Warehouse

10.1. OLAP

10.2. Herramientas OLAP

11. Big Data. Introducción al ecosistema de Hadoop

6. Cronograma

6.1. Cronograma de la asignatura *

Sem	Actividad presencial en aula	Actividad presencial en laboratorio	Otra actividad presencial	Actividades de evaluación
1	1 sistemas gestores de bases de datos Duración: 02:00 LM: Actividad del tipo Lección Magistral			
2	2 Almacenamiento de datos Duración: 02:00 LM: Actividad del tipo Lección Magistral			
3	3 Procesamiento de interrogaciones Duración: 02:00 LM: Actividad del tipo Lección Magistral			Tecnologías en el DW TG: Técnica del tipo Trabajo en Grupo Evaluación continua Duración: 15:00
4		4 Caso práctico: MySQL Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
5		4 Caso práctico: MySQL Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
6	5.1 Introducción a data warehouse Duración: 02:00 LM: Actividad del tipo Lección Magistral			Integración de procesos en el DW: ETLs TG: Técnica del tipo Trabajo en Grupo Evaluación continua Duración: 12:00
7	5.2 Arquitectura del data warehouse Duración: 02:00 LM: Actividad del tipo Lección Magistral			
8	5.3 Módulos del data warehouse Duración: 02:00 LM: Actividad del tipo Lección Magistral			
9	6.1 Funcionalidad ETL Duración: 02:00 LM: Actividad del tipo Lección Magistral			
10	7.1 Diseño multidimensional básico Duración: 02:00 LM: Actividad del tipo Lección Magistral			
11	7.2 Diseño multidimensional extendido Duración: 02:00 LM: Actividad del tipo Lección Magistral			
12	8.1 Metodologías de diseño del data warehouse Duración: 01:00 LM: Actividad del tipo Lección Magistral 8.2 Bus común del data warehouse Duración: 01:00 LM: Actividad del tipo Lección Magistral			Integración de Hadoop en los procesos de BI && DW TG: Técnica del tipo Trabajo en Grupo Evaluación continua Duración: 12:00

13	8.3 DW 2.0 Duración: 01:00 LM: Actividad del tipo Lección Magistral 8.4 Consulta de información del Data Warehouse Duración: 01:00 LM: Actividad del tipo Lección Magistral			
14	9 Infraestructuras de Big Data. Ecosistema de Hadoop Duración: 02:00 LM: Actividad del tipo Lección Magistral			Procesos de Implantación del DW: Análisis y Explotación del Data Warehouse TG: Técnica del tipo Trabajo en Grupo Evaluación continua Duración: 09:00
15	10 Data Streamming y otras problemáticas Duración: 02:00 LM: Actividad del tipo Lección Magistral			
16				
17				Examen EX: Técnica del tipo Examen Escrito Evaluación sólo prueba final Duración: 02:00 Proyecto práctico OT: Otras técnicas evaluativas Evaluación sólo prueba final Duración: 00:00 Examen Individual EX: Técnica del tipo Examen Escrito Evaluación continua Duración: 02:00

Las horas de actividades formativas no presenciales son aquellas que el estudiante debe dedicar al estudio o al trabajo personal.

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.

7. Actividades y criterios de evaluación

7.1. Actividades de evaluación de la asignatura

7.1.1. Evaluación continua

Sem.	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
3	Tecnologías en el DW	TG: Técnica del tipo Trabajo en Grupo	No Presencial	15:00	25%	3 / 10	Ce 12/16 Ce 13/18 CG-2/CE45
6	Integración de procesos en el DW: ETLs	TG: Técnica del tipo Trabajo en Grupo	No Presencial	12:00	25%	3 / 10	Ce 12/16 CG-2/CE45 Ce 13/18
12	Integración de Hadoop en los procesos de BI && DW	TG: Técnica del tipo Trabajo en Grupo	No Presencial	12:00	15%	3 / 10	Ce 13/18 CG-2/CE45 Ce 12/16
14	Procesos de Implantación del DW: Análisis y Explotación del Data Warehouse	TG: Técnica del tipo Trabajo en Grupo	No Presencial	09:00	10%	3 / 10	CG-2/CE45 Ce 12/16 Ce 13/18
17	Examen Individual	EX: Técnica del tipo Examen Escrito	Presencial	02:00	25%	4 / 10	CG-2/CE45 Ce 12/16 Ce 13/18 Ce 44

7.1.2. Evaluación sólo prueba final

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
17	Examen	EX: Técnica del tipo Examen Escrito	Presencial	02:00	80%	4 / 10	Ce 13/18 Ce 12/16
17	Proyecto práctico	OT: Otras técnicas evaluativas	No Presencial	00:00	20%	3.5 / 10	CG-2/CE45 Ce 12/16

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	Ce 12/16 CG-2/CE45 Ce 13/18 Ce 44

7.2. Criterios de evaluación

La asignatura de Arquitectura del Data Warehouse se evaluará mediante 4 proyectos:

- tres cubren específicamente el desarrollo de un proyecto de implantación de un Data Warehouse
- uno se enfoca al análisis de alguna de las diversas tecnologías partícipes en la Arquitectura de un Data Warehouse

Los proyectos se realizarán en grupos de **hasta 4 alumnos** de entre los matriculados de la asignatura al inicio del curso.

Para poder superar la asignatura, en la **convocatoria de junio**, se establecen los siguientes requisitos:

1. Obtener un mínimo de 50 puntos sobre los 100 disponible en el cómputo global de la evaluación sumativa
2. Es OBLIGATORIO realizar todos los proyectos
3. En los proyectos se debe obtener una nota mínima igual o superior al 30% de la valoración del mismo (ver tabla de valoración sumativa)
4. Es OBLIGATORIO superar una prueba de evaluación individual

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

1. No habrá proyectos, sólo se realizará un examen que cubrirá los aspectos teóricos y prácticos de la asignatura
2. Obtener un mínimo de 50 puntos sobre los 100 disponibles en el cómputo global

Para aquellos alumnos que de forma extraordinaria, no puedan realizar la evaluación continua, y previa petición por escrito durante los primeros 15 días del curso, la forma de evaluación de la asignatura será la siguiente, **siendo excluyente con la evaluación continua**.

1. Examen en junio en la fecha establecida en el calendario oficial de exámenes por jefatura de estudios.
2. Realización de un proyecto práctico, en grupos de hasta cuatro alumnos, compuesto por tres entregas, valorado en 20 puntos.

Para poder superar la asignatura en la **convocatoria de junio (mediante la evaluación por prueba final)**, se establecen los siguientes requisitos:

1. La valoración de examen es de 80 puntos.
2. Para poder aprobar la asignatura en esta convocatoria es necesario obtener una nota mínima igual o superior al 40% de la valoración del examen.
3. En el proyecto práctico es necesario obtener una valoración mínima del 35% de la valoración del mismo.
4. Obtener un mínimo de 50 puntos sobre los 100 disponibles en el cómputo global (examen + proyecto práctico)

A los alumnos que opten por este sistema de evaluación extraordinario en la convocatoria de junio y no superen la asignatura **no se les guardará ninguna de la calificaciones** para la convocatoria de julio y sucesivas.

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	