



Bases de Datos y Sistemas de Información

Grado en Ingeniería Informática

Unidad Didáctica 1: Bases de Datos Relacionales

Parte 1: Conceptos básicos

(Doc. UD1.1)

Curso 2020/2021



El contenido de este documento no debe considerarse una aportación novedosa ya que se ha realizado utilizando casi literalmente el libro:

Celma, M.; Casamayor, J. C.; Mota, L.; Bases de datos relacionales. Pearson, Prentice Hall, 2003

El permiso explícito de los tres autores ha permitido su realización con el objetivo de facilitar el estudio de la materia de bases de datos para los alumnos de la ETSInf.

Índice

1 Introducción a las bases de datos	1
2 Sistema de información	1
3 Base de datos	2
4 Sistema de gestión de bases de datos	2
5 Características de las técnicas de bases de datos	2
6 Ejemplo	3
6.1 Descripción del sistema de información	3
6.2 Construcción del sistema de información	7



1 INTRODUCCIÓN A LAS BASES DE DATOS

El tratamiento automático de la información en un computador ha experimentado una profunda evolución desde los pioneros sistemas de gestión de ficheros hasta los actuales sistemas de gestión de bases de datos. En la actualidad, las técnicas de bases de datos representan la tecnología informática disponible para la organización y gestión de grandes volúmenes de datos, ocupando un lugar relevante en todas las áreas en las que son utilizados los computadores: gestión, ingeniería, ciencia, medicina, etc. Esta tecnología se caracteriza por:

- Dar servicio a usuarios con distintas necesidades.
- Gestionar grandes volúmenes de información.
- Asegurar la persistencia de los datos en el tiempo.

En este documento se introducen los conceptos fundamentales relacionados con la tecnología de bases de datos: sistema de información, base de datos y sistema de gestión de bases de datos. Aunque estos conceptos serán desarrollados a lo largo de la asignatura, esta introducción tiene como objetivo fijar desde el principio una terminología clara y unas ideas básicas que irán adquiriendo mayor significado a medida que se avance en el estudio de la materia.

2 SISTEMA DE INFORMACIÓN

Uno de los pilares de cualquier organización es la información que necesita para su funcionamiento; asimismo, una de sus actividades principales es el tratamiento de dicha información el cual, ya se haga de forma manual o automática, tiene como objetivo proporcionar a las personas autorizadas la información que necesitan en el momento y el lugar adecuados. Por ello, uno de los componentes básicos de cualquier organización es su sistema de información.

Un sistema de información (SI) es un conjunto de elementos ordenadamente relacionados entre sí de acuerdo a ciertas reglas, que aportan a la organización a la que sirven la información necesaria para el cumplimiento de sus fines.

Un sistema de información *informatizado* es aquél que está soportado por un computador; a partir de este punto, con la expresión "sistema de información" siempre se hará referencia a un sistema de información informatizado.

Las funciones básicas de un sistema de información son:

- Recogida de datos.
- Procesamiento y almacenamiento de datos.
- Elaboración y presentación de datos.

Para el cumplimiento de estas funciones, un sistema de información se compone de los siguientes elementos:

- Contenido: los datos, su descripción y los programas que los manipulan.
- Equipo físico: el computador que soporta el sistema de información.
- Equipo lógico: el sistema de comunicaciones, el sistema de gestión de base de datos, el sistema operativo, etc.
- Administrador: la persona o el equipo de personas responsables de asegurar la calidad y la disponibilidad de los datos.
- Usuarios.

De todos estos elementos, los datos constituyen la parte esencial del sistema, y son los que justifican su existencia. Para organizar y gestionar estos datos en el computador, se han desarrollado técnicas cuya evolución ha estado determinada, principalmente, por el desarrollo de la tecnología de los computadores,



así como por los nuevos requisitos y necesidades planteadas por los usuarios.

En la actualidad, las técnicas de bases de datos representan la tecnología informática disponible para la organización y gestión de grandes volúmenes de datos en un computador. Se puede afirmar que el núcleo de todo sistema de información actual es una base de datos, y que el diseño y creación de ésta constituyen una etapa importante en la construcción del sistema.

3 BASE DE DATOS

Una base de datos (BD) es una colección estructurada de datos.

En esta colección, los datos deben estar estructurados de forma que reflejen fielmente los objetos, las relaciones y las restricciones existentes en la parcela del mundo real representada por la base de datos (propiedades estáticas). Asimismo, y para que esta representación sea fiable, la base de datos debe ser sensible a los sucesos del mundo real y debe evolucionar para reflejar los cambios que estos sucesos puedan provocar en la parcela del mundo representada (propiedades dinámicas). Los mecanismos de estructuración de datos que se pueden utilizar dependen del sistema informático con el que se vaya a crear y manipular la base de datos: lo que se conoce como sistema de gestión de bases de datos.

4 SISTEMA DE GESTIÓN DE BASES DE DATOS

Un sistema de gestión de bases de datos (SGBD) es una herramienta software (conjunto de programas) que permite la creación y manipulación de bases de datos.

Es importante destacar que un sistema de gestión de bases de datos es una herramienta de propósito general que permite crear bases de datos de cualquier tamaño y complejidad y con propósitos específicos distintos.

Actualmente existen SGBD libres como pueden ser el *PostgreSQL* o el *MySQL* y de pago entre los que destacan el *ORACLE* o el *ACCESS* (de Microsoft) como sistemas comerciales más extendidos.

Todo sistema de gestión de bases de datos está basado en un modelo de datos. Los modelos de datos proporcionan estructuras de datos predefinidas con sus operadores asociados. De una forma general, un modelo de datos se puede definir como sigue:

Un modelo de datos es un conjunto de conceptos y reglas que permiten estructurar los datos resultantes de la observación de la realidad, de forma que queden representadas todas sus propiedades, tanto estáticas como dinámicas.

Los sistemas de gestión de bases de datos más utilizados actualmente están basados en el *modelo* relacional de datos que se estudiará en la parte 2 de la unidad didáctica 1 (documento UD1.2). SGBD más antiguos estaban basados en el *modelo jerárquico de datos* y en el *modelo red*.

5 CARACTERÍSTICAS DE LAS TÉCNICAS DE BASES DE DATOS

Aunque una base de datos es una colección estructurada de datos, no cualquier conjunto de estructuras de datos se puede considerar una base de datos. La tecnología de bases de datos se ha desarrollado intentando dar respuesta a las crecientes exigencias de funcionalidad y eficiencia que los usuarios plantean a los sistemas de información. Las características que definen esta tecnología, y que la diferencian de otras técnicas de gestión de datos anteriores, son las siguientes:

Integración de toda la información de la organización. La base de datos se crea para dar servicio
a toda o a una parte importante de la organización, y no para unos usuarios particulares; de esta
forma se evita la redundancia de datos dentro del sistema de información y los problemas de
inconsistencia derivados de ella.



- *Persistencia de los datos*. Los datos deben estar disponibles en todo momento, lo que significa que la base de datos debe almacenarse en un dispositivo de memoria secundaria.
- Accesibilidad simultánea para distintos usuarios. Debido al carácter integrador que tiene la base de datos, ésta tendrá que ser compartida por distintos grupos de usuarios, lo que significa que éstos podrán acceder simultáneamente a los datos.
- Descripción unificada de los datos e independiente de los programas. Las estructuras de datos que constituyen la base de datos se definen independientemente de los programas (o aplicaciones) que las van a manipular, quedando almacenada esta definición junto a los datos para uso del SGBD.
- Independencia de los programas respecto a la representación física de los datos. Las aplicaciones que se desarrollen para manipular los datos deben ser independientes de la implementación elegida para las estructuras de la base de datos. A esta característica se conoce como independencia de datos.
- Definición de vistas parciales de los datos para distintos usuarios. Debido también al carácter integrador de la base de datos, en ésta se recogen los datos que interesan a cada grupo de usuarios de la organización, con lo que se incrementa su tamaño y complejidad. Para que los usuarios puedan seguir teniendo su visión particular de los datos sin que se pierda este carácter integrador, el SGBD debe permitir definir vistas parciales de la base de datos que contengan sólo aquellos datos que son relevantes para cada uno de los grupos.
- Mecanismos para controlar la integridad y la seguridad de los datos. Para que la base de datos refleje fielmente la realidad de la cual es una representación, el SGBD debe asegurar en todo momento la calidad de la información almacenada (integridad), evitando que ésta se deteriore por un uso incorrecto (actualizaciones que no son válidas, accesos concurrentes no controlados, etc.). Asimismo, debe asegurar que a la información almacenada sólo acceden las personas autorizadas y en la forma autorizada (seguridad).

Estas características se pueden resumir diciendo que las técnicas de bases de datos se han desarrollado con el objetivo de *integrar la información del sistema para evitar redundancias, sin que por ello se pierdan las distintas perspectivas que de ella tienen los usuarios*, y que las herramientas de software (SGBD) que se construyen para aplicar estas técnicas deben asegurar la *independencia, la integridad y la seguridad de los datos*.

Para cumplir con estos objetivos, los sistemas de gestión de bases de datos, independientemente del modelo en que se basen y del fabricante responsable, disponen de componentes con funciones bien definidas que permiten definir y manipular una base de datos satisfaciendo todos estos requisitos.

6 EJEMPLO

Para ilustrar de forma sencilla los conceptos e ideas introducidos en este capítulo, en este apartado se va a presentar un sistema de información simplificado relativo a la gestión docente en una universidad española.

6.1 Descripción del sistema de información

El Vicerrectorado de Ordenación Académica de una universidad ha decidido crear un sistema de información para la gestión del Plan de Ordenación Docente (POD) del curso académico.

El POD se diseña al principio de cada curso académico, y contiene la planificación de la docencia de todos los estudios que se imparten en la universidad durante el curso.

Algunos datos sobre la estructura de la universidad española que son relevantes para la construcción de este sistema son:

 El contenido de los estudios conducentes a la obtención de un título académico se define en el plan de estudios de la titulación. Los estudios se organizan en asignaturas (materias) valoradas con un número de créditos teóricos (de aula) y prácticos (de laboratorio), lo que determina el número de horas lectivas de la asignatura.



- Las asignaturas de un plan de estudios se organizan en cursos académicos (año académico) divididos en semestres.
- Para el cumplimiento de sus funciones, la universidad se organiza en centros y departamentos.
- Los centros son responsables de organizar la docencia de los estudios de una o varias titulaciones, lo que significa realizar las siguientes tareas: definir los grupos (de teoría y de prácticas) de cada asignatura, organizar los horarios, gestionar la matrícula de los alumnos, distribuir los alumnos por grupos, etc. Por ejemplo, la ETSInf es responsable de organizar la enseñanza de los estudios del Grado en Ingeniería Informática.
- Los departamentos son los responsables directos de impartir la docencia; en ellos se agrupan los
 profesores pertenecientes a varias materias afines. Algunos departamentos de la universidad
 son: Matemática Aplicada (DMA), Física Aplicada (DFA), Sistemas Informáticos y Computación
 (DSIC), etc. Cada departamento tiene asignada la docencia de asignaturas relacionadas con sus
 áreas de conocimiento; estas asignaturas pueden pertenecer al plan de estudios de cualquier
 titulación de la universidad. Por ejemplo, el DSIC imparte la docencia de las asignaturas de
 programación en todos los centros de la universidad.

Antes de esta iniciativa del vicerrectorado, cada centro y cada departamento disponía de su propio sistema de información independiente; por ello, y de acuerdo al carácter integrador de la tecnología de bases de datos, el sistema que se cree, concretamente la base de datos que se diseñe, deberá integrar toda la información sobre la docencia en la universidad, permitiendo al mismo tiempo que los centros y los departamentos sigan teniendo la perspectiva de los datos que les resulte más cómoda para el cumplimiento de sus funciones. A continuación se presentan las perspectivas de centros y departamentos.

Perspectiva de un centro

Desde el punto de vista de un centro, la información que necesita para realizar sus funciones es la relacionada con las asignaturas de los planes de estudios de las titulaciones que imparte. De cada una de estas asignaturas necesita conocer: el código, el nombre, el curso y semestre en el que se imparte, el departamento encargado de su docencia, el número de créditos de teoría y de prácticas asignados en el plan de estudios, el número de grupos de teoría y de prácticas creados para el presente curso académico, los profesores encargados de la docencia de la asignatura, y, para cada uno de ellos, el número de grupos de teoría y de prácticas asignados en ella. Para simplificar, se omite la distribución de los profesores de una asignatura en los grupos creados. Toda esta información queda recogida en los informes que con más frecuencia maneja el centro (en las Figuras 1.1, 1.2 y 1.3 se presentan fragmentos de informes de la ETSInf.

06 de septiembre de	2012 Profesore	es de la ETSInf	Página 1 de 4
Departamento	Código profesor	Nombre	Teléfono
DMA	LBP	Bos Pérez, Luisa	3545
	JCP	Cerdá Pérez, Juan	3222
	PMG	Martí García, Pedro	3412
DFA	MRC	Ruiz Cantó, María	3675
L			

Figura 1.1: Fragmento de listado de profesores que imparten docencia en la ETSInf



06 de se	eptiembre de 2012 Asignatura	gnaturas de la ETSInf			de 8
	Grado en Ing	eniería In	formática		
C/S	Asignatura	Créditos teoría	Créditos prácticas		
1A	Análisis Matemático	11545	DMA	4,5	1,5
	Matemática Discreta	11547	DMA	4,5	1,5
	Fundamentos Físicos Informática	11540	DFA	4,5	1,5
	Intr. a la Informática y Programación	11541	DSIC	4,5	1,5
	Fundamentos de Computadores	11542	DISCA	4,5	1,5
1B	Estadística	11538	DEIOAC	4,5	1,5
L	···				

Figura 1.2: Fragmento del listado de asignaturas de la ETSInf

06 de sep	06 de septiembre de 2012 Docencia por semestres y asignaturas en la ETSInf						
C/S	Asignatura	Grupos teoría	Grupos prácticas	Profesor	Grupos teoría	Grupos prácticas	
1A	11545	2	4	Cerdá Pérez, Juan	1	2	
				Martí García, Pedro	1	2	
	11547	2	4	Bos Pérez, Luisa	1	2	
				Cerdá Pérez, Juan	1	2	

Figura 1.3: Fragmento del listado de la docencia por semestres

Perspectiva de un departamento

La información que necesita un departamento para realizar sus funciones es la relacionada con las asignaturas de cuya docencia es responsable, así como la relacionada con los profesores que son miembros del departamento. De cada asignatura necesita conocer: el código, el nombre, el curso y semestre, la titulación a la que pertenece, el centro en el que se imparte, el número de créditos de teoría y de prácticas, la cantidad de grupos de teoría y de prácticas creados para el presente curso académico, y los profesores del departamento asignados a la asignatura, con el número de grupos a su cargo. De cada profesor necesita conocer: el nombre, el teléfono, la categoría y las asignaturas asignadas con el número de grupos que debe impartir. Esta información queda recogida en los informes que con más frecuencia maneja el departamento (en las Figuras 1.4, 1.5, 1.6 y 1.7 se presentan fragmentos de informes del Departamento de Matemática Aplicada).

DMA:	Listado de profesores 06/09/2011				
Código	Nombre	Categoría	Teléfono		
LBP	Bos Pérez, Luisa	Titular	3545		
JCP	Cerdá Pérez, Juana	Titular	3222		
PMG	Martí García, Pedro	Titular	3412		

Figura 1.4: Fragmento del listado de profesores del DMA



DMA			Listado de asignaturas			06/09/12
Título	Centro	C/S	Asignatura	Código	Teoría	Prácticas
GII	ETSInf	1A	Análisis Matemático	11545	4,5	1,5
			Matemática Discreta	11547	4,5	1,5
		1B	Álgebra	11546	4,5	1,5

Figura 1.5: Fragmento del listado de asignaturas del DMA

	DMA:	Docencia por asignaturas						6/09/12
C/S	Título	Centro	Asig	Grupos teoría	Grupos prácticas			Grupos prácticas
1A	GII	ETSInf	11545	2	4	Cerdá Pérez, Juana	1	2
						Martí García, Pedro	1	2
			11547	2	4	Bos Pérez, Luisa	1	2
						Cerdá Pérez, Juana	1	2
			• • • •			•••		

Figura 1.6: Fragmento del listado de la docencia del DMA por asignaturas

DMA:	Docencia	por pi	06	6/09/12		
Profesor	Asignatura	Título	Centro	C/S		Grupos prácticas
Bos Pérez, Luisa	11547	GII	ETSInf	1A	1	2
					Total créd	ditos: 7,5
Cerdá Pérez, Juan	a 11545	GII	ETSInf	1A	1	2
	11547	GII	ETSinf	1A	1	2
					Total cré	ditos: 15
Martí García, Pedro	o 11545	GII	ETSInf	1A	1	2

Figura 1.7: Fragmento del listado de la docencia del DMA por profesores

Con independencia de la tecnología de gestión de datos que haya utilizado cada centro y cada departamento para construir su sistema de información, de los informes anteriores se desprende la existencia de *redundancia de datos* dentro de la universidad. Por ejemplo:

- Los datos de un profesor se repiten en cada centro en el que el profesor imparte clases y en el departamento al que pertenece.
- Los datos de una asignatura se repiten en el departamento encargado de su docencia y en el centro en el que se imparte el título al que pertenece la asignatura.
- Los datos sobre el POD del año académico (es decir, la asignación de la docencia de las asignaturas a los profesores) se repiten en el departamento y en el centro en el que se imparte el título al que pertenece la asignatura.

La redundancia de datos puede generar *inconsistencia* dentro del sistema de información; por ejemplo, hay un profesor del departamento de DMA con docencia en la ETSInf cuyo nombre es distinto en el centro y en el departamento (¿se llama Juan o Juana?). Asimismo, la redundancia de datos provoca que los procesos de actualización sean más costosos; por ejemplo, un cambio en el POD de una asignatura debe realizarse tanto en el departamento como en el centro correspondiente.



6.2 Construcción del sistema de información

La creación de un sistema de información único que integre toda la información de la universidad relativa a la docencia, que dé servicio a los centros y departamentos según sus necesidades y que sea seguro y fiable, obliga a utilizar la tecnología de bases de datos.

La creación de este sistema se inicia con el análisis de las necesidades de los potenciales usuarios (centros y departamentos), necesidades que de forma muy simplificada se han comentado en el apartado anterior. Este análisis permite conocer la información (datos) y operaciones (consultas, informes, etc.) que con más frecuencia van a realizar los usuarios.

A continuación, y dependiendo del sistema de gestión de bases de datos que se vaya a utilizar como soporte del sistema de información, se debe pasar a definir el *esquema lógico* de la base de datos. Esto significa definir un conjunto de estructuras de datos en las cuales se puedan organizar los datos recogidos en la etapa de análisis, de forma que se capture lo mejor posible el conocimiento que se tiene de la realidad que se desea representar. En un sistema relacional, el esquema lógico consistiría en un conjunto de *relaciones* (tablas con filas y columnas) de la forma siguiente:¹

Relación Departamento

código	go nombre directo		teléfono						
DMA	Matemática Aplicada	Juan García	3570						
DFA	Física Aplicada	José Ruíz	3540						

Relación Título

código	nombre	centro
GIGT	Grado de Ingeniero en Geomática y Topografía	ETSIGCT
GII	Grado en Ingeniería Informática	ETSInf

Relación Centro

código	nombre	director	teléfono
ETSIGCT	ETSI Geodésica, Cartográfica y Topográfica	Pedro Ruíz	3578
ETSInf	ETSI Informática	José Esteban	3776
		•••	

Relación Profesor

código	nombre	teléfono	categoría	dpto				
JCP	Juana Cerdá Pérez	3222	Titular	DMA				
PMG	Pedro Martí García	3412	Titular	DMA				
LPB	Luisa Bos Pérez	3545	Titular	DMA				

 $^{^{}m 1}$ En el documento UD1.2 se presentará con profundidad el Modelo Relacional de Datos.

_



Relación Asignatura

código	nombre	semestre	Т	Р	GT	GP	título	dpto
11545	Análisis Matemático	1A	4 ,5	1,5	2	4	GII	DMA
11547	Matemática Discreta	1A	4 ,5	1 ,5	2	4	GII	DMA
11546	Álgebra	1B	4 ,5	1 ,5	1	3	GII	DMA

Relación Docencia

cod_pro	cod_asg	GT	GP	
JCP	11545	1	2	
JCP	11547	1	2	
LBP	11547	1	2	
PMG	11545	1	2	

donde el significado de algunas columnas es el siguiente:

- En la relación *Profesor*:
 - La columna dpto representa el código del departamento al que pertenece el profesor.
- En la relación Asignatura:
 - La columna semestre representa el curso y el semestre en los que se imparte la asignatura (por ejemplo, el valor '1B', significa 1er curso 2º semestre).
 - La columna T representa los créditos teóricos asignados en el plan de estudios a la asignatura.
 - La columna *P* representa los créditos prácticos asignados en el plan de estudios a la asignatura.
 - La columna *GT* representa el número de grupos de teoría creados para la asignatura en el presente curso académico.
 - La columna *GP* representa el número de grupos de prácticas creados para la asignatura en el presente curso académico.
 - La columna Dpto representa el código del departamento al que está adscrita la asignatura.
- Cada fila de la relación Docencia representa la siguiente información: el profesor de código Cod_pro imparte GT grupos de teoría y GP grupos de prácticas de la asignatura de código Cod_asg.

Este conjunto de relaciones (tablas) recoge de forma estructurada todos los datos necesarios para satisfacer las necesidades de los centros y de los departamentos, aunque se ha diseñado sin pensar en ninguna aplicación (consulta, informe, etc.) particular; es decir, se ha hecho una descripción unificada de los datos independiente de las aplicaciones actuales o futuras que vayan a acceder a ellos.

Evidentemente, al integrar los datos de distintos usuarios se ha perdido la perspectiva original que éstos tenían de los mismos. Un buen sistema de gestión de bases de datos debe permitir definir, a partir del esquema lógico, vistas parciales de los datos (esquemas externos) para cada grupo de usuarios particular en las que se recojan los datos que son relevantes para el cumplimiento de las funciones de dichos usuarios. En el ejemplo se debería definir un esquema externo para cada centro y un esquema externo para cada



departamento de la universidad; en la Figura 1.8 se presenta el esquema externo correspondiente al departamento DMA; el esquema consiste en un conjunto de relaciones *derivadas*, definidas a partir de las relaciones del esquema lógico, en las que se incluyen sólo los datos que el departamento necesita para obtener los informes que se han mostrado anteriormente. Estas relaciones derivadas son virtuales: sus datos no se almacenan explícitamente y sirven para ocultar a los usuarios los datos que no necesitan.

La definición de la base de datos en tres niveles, lógico (definición de relaciones), externo (definición de vistas parciales) y físico (cómo se almacenan los datos en disco), favorece la *independencia de datos*, que asegura que los cambios en un esquema no afectarán a los esquemas superiores. Así, la independencia entre el nivel lógico y el nivel físico permite hacer cambios en la implementación de cualquiera de las relaciones sin que ello afecte al esquema lógico ni, consecuentemente, a los esquemas externos, consiguiendo de esta forma que los programas sean independientes de los cambios en la organización física de los datos. Por otro lado, la independencia entre el nivel lógico y el nivel externo permite hacer algunos cambios en el esquema lógico que no afectarán a los programas; por ejemplo, se pueden incluir más atributos (columnas) en la relación *Profesor* sin que ello afecte a los esquemas externos, ni por lo tanto a los programas; en general, cualquier cambio en el esquema lógico que permita definir las mismas relaciones derivadas que aparecen en los esquemas externos será totalmente transparente para los programas. Como conclusión, se puede afirmar que la independencia de datos representa la posibilidad de construir aplicaciones más seguras y duraderas, reduciendo significativamente el coste del mantenimiento del software.

Por último, y teniendo en cuenta el carácter integrador de la base de datos diseñada, será imprescindible que el SGBD disponga de mecanismos para controlar el acceso simultáneo de varios usuarios a los mismos datos, situación que sucederá con frecuencia; también, y en la línea de conseguir sistemas de información cada vez más seguros, el SGBD deberá disponer de herramientas para recuperar la base de datos frente a cualquier fallo del usuario o del sistema.



Esquema lógico

Departamento				Centro					
	código	nombre	director	teléfono	código	nombre	director	teléfono	
		• • • •							

Profesor

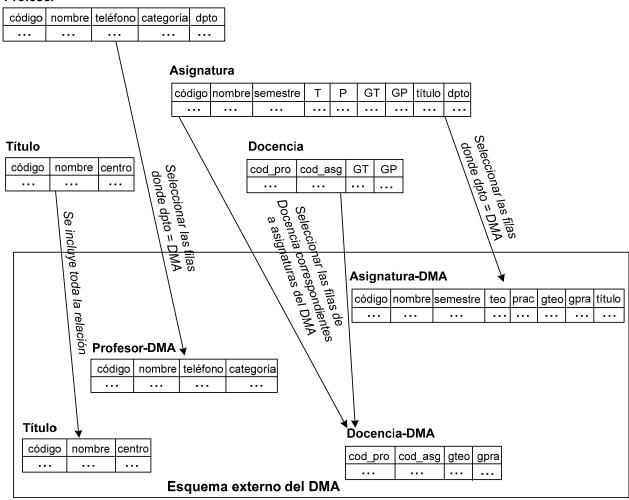


Figura 1.8 Definición del esquema externo del departamento de código DMA