

NOMBRE DE LA ASIGNATURA:	Bases de Datos
DEPARTAMENTO	Ingeniería de Sistemas Telemáticos
NOMBRE (EN INGLÉS):	Databases
MATERIA QUE DESARROLLA (BOE):	Ingeniería de Sistemas Informáticos
DESCRIPCIÓN (BOE):	Técnicas, tecnologías y sistemas en campos tales como: sistemas operativos, bases de datos, arquitectura de sistemas informáticos, arquitectura de sistemas distribuidos, sistemas inteligentes, ingeniería de sistemas complejos.
ÁREAS DE CONOCIMIENTO	Arquitectura y Tecnología de Computadores, Ingeniería Telemática, Lenguajes y Sistemas Informáticos, Ciencias de la Computación e Inteligencia Artificial
NOMBRE DE LA COORDINADORA:	Dra. Carmen Costilla Rodríguez
NOMBRE DE PROFESORA/GRUPO	Dra. Carmen Costilla Rodríguez / Grupo 5º. Op.
ASIGNATURA APROBABLE POR CURSO	Sí
NÚMERO DE EXÁMENES PARCIALES	Dos
EXISTENCIA DE PRUEBAS DE EVALUACIÓN CONTINUA	Sí
PERIODOS ESTIMADOS DE REALIZACIÓN DE PARCIALES Y PRUEBAS	Presentación de trabajos: de Noviembre a Enero, se fija sobre la marcha. 1er. Parcial: 1ª quincena de Diciembre 2º Parcial: unos días antes del examen final
TIPO DE EXAMEN FINAL	Examen escrito, con una parte teórica (sin libros) y otra práctica (con libros)



### Conocimientos previos recomendados

- **Necesario.** Programación en lenguaje Orientado a Objetos,
- **Recomendable.** Sistemas Operativos Distribuidos y Software de Comunicaciones.

### Objetivos Docentes

El estudio de los siguientes conceptos, técnicas y herramientas actuales de bases de datos:

- **Conceptos.** Modelado Conceptual de la Información. **Modelos de Datos:** 'Entity-Relationship', Relacional, Objeto-Relacional y XML. **Teorías:** Diseño Relacional. Concepto de Transacción, Concurrencia y Serializabilidad.
- **Técnicas.** Estándares y modelos de referencia: Arquitecturas de los Sistemas de Gestión de Bases de Datos, SGBD. Lenguaje SQL:1999, SQL:2003 (objeto-relacional) y SQL/XML:2006. Procesamiento de consultas. Control de concurrencia y gestión de transacciones. Persistencia y recuperación frente a fallos. Aspectos distribuidos: acceso remoto a BDs, Cliente/Servidor, Web Server y BD Distribuidas. Interoperabilidad entre bases de datos heterogéneas. Integración virtual, flexible y dinámica de fuentes de datos Web heterogéneas en los Sistemas de Información Web. Web Semántica y lenguaje SPARQL (2008).
- **Herramientas.** Tecnología de Bases de Datos y herramientas asociadas.
- **Prácticas.** El programa está repleto de pequeños ejercicios. Se realizan algunas prácticas de Laboratorio y se ofertan trabajos en grupo. Además, se puede practicar con ORACLE 10g, SQL Server y ADO.NET en bases de datos xml-objeto-relacionales: diseño, SQL, APIs y pequeños ejemplos con JSP y con SPARQL.

### Programa Docente

Tema I. Introducción a Bases de Datos. Arquitectura de los Sistemas de Gestión de Bases de Datos, Estándar ANSI/X3/SPARC.

Tema II. Modelado Conceptual y Modelos de Datos Semánticos.

II.1. Modelado Conceptual. Introducción a los Modelos de Datos.

II.2. Modelo de Datos 'Entity-Relationship'. Definiciones y conceptos. Ejemplos y ejercicios.

Tema III. Bases de Datos Relacionales.

III.1. Introducción a los Modelos de Datos clásicos: Relacional, Codasyl y Jerárquico.

III.2. Sistemas y Lenguajes de Bases de Datos clásicas: Relacional, Codasyl y Jerárquicas.

III.3. Modelo de Datos Relacional. Definiciones y Conceptos.

III.4. Lenguajes Formales del Modelo de Datos Relacional: Álgebra y Cálculo Relacional.

Tema IV. Diseño de Bases de Datos Relacionales.

IV.1. Introducción al Diseño de las Bases de Datos Relacionales.

IV.2. Introducción a la Teoría Relacional.

IV.3. Dependencias Funcionales.

IV.4. Forma Normal de la Relación. Ejemplos y Ejercicios

Tema V. Lenguaje SQL de Bases de Datos Objeto-Relacionales, BDOR, y XML.

V.1. Introducción a los Lenguajes de la Tecnología Relacional.

V.2. SQL. Lenguaje de Definición de Datos, DDL. Sintaxis en BNF.

V.3. Pre-compilación y ejecución de un módulo SQL embebido en un lenguaje anfitrión (*host*).

V.4. SQL. Lenguaje de Manipulación de Datos, DML. Sintaxis en BNF.

V.5. SQL. Lenguaje de Definición de Vistas. Ejemplos.

V.6. Expresiones de Consultas en SQL. Gramática SQL en BNF. Ejemplos y ejercicios.

V.7. Estándares SQL:92 y SQL:99

V.8. SQL:2003. Modelo de Datos SQL Objeto-Relacional y SQL/XML. Tecnología actual de Bases de Datos.

V.9. Estándar SQL/XML:2006 y XML. Ejemplos y Ejercicios.



**Tema VI. Interoperabilidad entre Múltiples Bases de Datos Heterogéneas y Distribuidas.**

- VI.1. Concepto de Interoperabilidad entre Sistemas de Información Heterogéneos.
- VI.2. Interoperabilidad entre Sistemas de Bases de Datos: RDA, ODBC, JDBC y C/S (net, open y star).
- VI.3. Sistemas de Bases de Datos Múltiples y Heterogéneos. Autonomía, Heterogeneidad y Distribución.
- VI.4. Organización de BD Relacionales Distribuidas, Homogéneas y Altamente Integradas.
- VI.5. Bases de Datos Web. Arquitectura *Mediador-Wrapper*.
- VI.6. La Web profunda con Bases de Datos. Web Semántica y SPARQL (2008).

**Tema VII. Arquitectura de los SGBD Distribuidos.**

- VII.1. Niveles de Transparencia en los Sistemas de Bases de Datos Distribuidos, SBDD.
- VII.2. Estándares y Modelos Arquitecturales de los SBDD.
- VII.3. Arquitecturas Cliente/Servidor (dos capas) y Web (tres o más capas).
- VII.4. Datos semi-estructurados y Sistemas para consultar fuentes de datos Web heterogéneas.

**Tema VIII. Diseño de Bases de Datos Distribuidas.**

- VIII.1. Alternativas de Diseño y Aspectos de la Distribución de los Datos.
- VIII.2. Algoritmos de Fragmentación. Programas de localización de datos.

**Tema IX. Control Semántico de los Datos.**

- IX.1. Gestión de Vistas. Seguridad de los Datos.
- IX.2. Control de la Integridad Semántica de Datos.
- IX.3. Objetos relevantes para la semántica de la BDOR y XML con SQL:2006.
- IX.4. Codificación de Triggers y otras técnicas del Control Semántico. Ejemplos y ejercicios.

**Tema X. Procesamiento de Consultas Distribuidas.**

- X.1. Objetivos y Niveles del Procesamiento de Consultas.
- X.2. Descomposición de la Consulta Global Distribuida y Localización de Datos Locales.

**Tema XI. Teoría de Concurrencia. Gestión de Transacciones Distribuidas y Control de Concurrencia.**

- XI.1. Introducción a la Concurrencia. El problema de la interferencia.
- XI.2. Definición, Propiedades y Tipos de Transacciones.
- XI.3. Teoría de la Serializabilidad.
- XI.4. Control de Concurrencia, técnicas y algoritmos: *Locking, Timestamping, Enfoques Optimista y Pesimista, Gestión de Bloqueos*.

**Tema XII. Persistencia y Fiabilidad de los Sistemas de Bases de Datos Distribuidos.**

- XII.1. Concepto de Persistencia y Medida de Fiabilidad.
- XII.2. Gestor de Datos. Mecanismos de persistencia y Recuperación frente a fallos: *Logs y Buffers, Chekpoint, Undo/Redo, etc.*
- XII.3. Protocolos de Fiabilidad Centralizada y Distribuida (*Two-phase Commit*).

**Bibliografía**

**Libros de Texto:**

1. **Sistemas de Bases de Datos. Conceptos, Técnicas y Lenguajes**, C. Costilla, S. Publicaciones E.T.S. Ingenieros de Telecomunicación-UPM, ISBN: 84-7402-271-1, 465 páginas, 1999
2. **Principles of Distributed Database Systems**, M. Tamer Özsu and P. Valduriez, Prentice-Hall, 2<sup>nd</sup> edition, ISBN: 0-13-607938-5, 666 pages, 1999.
3. **Apuntes complementarios del programa** en el sitio Web <http://sinbad.dit.upm.es>

**Evaluación Continua:** Para los que asisten a clase existe la opción de realizar trabajos y dos parciales, lo que permite aprobar la asignatura por curso.

**Examen Final:** Examen con una parte teórica (sin libros) y otra práctica (con libros).

**Profesorado:** Coordinadora y Profesora: Dra. Carmen Costilla Rodríguez

Home page: <http://sinbad.dit.upm.es>

e-mail: [costilla@dit.upm.es](mailto:costilla@dit.upm.es), Despacho: C-216, Laboratorio: B-205/1

Horario: Jueves de 18 a 20 horas y Viernes de 16 a 18 horas.

Tutorías: 2 horas después del horario de clases.



## Databases (BSDT 5403), <http://sinbad.dit.upm.es> Teaching Goals, Programme and Methodology

Credits: 6, 5th academic course, 1st. semester, academic year: **2010-11**  
Time-table: Thursday at 18:00 to 20:00 h. and Friday at 16:00 to 18:00 h.

### *Previous recommended knowledge*

- **Necessary.** Some practice about Object-Oriented Programming Language,
- **Recommendation.** Distributed Operating Systems and Communication Software.

### *Teaching Goals*

The main interest of this course could be summarized in the next database topics:

- **Concepts:** Conceptual Modelling Abstractions and the following data models: Entity-Relationship, Relational, Object-Relational and XML. Theories: Relational Design, the Transaction concept and the Serializability.
- **Techniques:** Standards and reference models: DBMS Architecture (ANSI/X3/SPARC), SQL:1999, SQL:2003 object-relational language and SQL/XML:2006. Query processing, concurrency control, transaction management and recovery in database systems. Other distributed and interoperable aspects are described: Remote Database Access, Client/Server and Web Server architectures, Web databases and Distributed databases. The Semantic Web and SPARQL are shortly introduced in order to virtually integrate heterogeneous and distributed Web data sources.
- **Tools:** Some technological tools of database area are described.
- **Practices:** The programme is plenty of brief exercises. Some laboratory practices can be done and some teamwork is offered. You can download ORACLE 10g, SQL Server y ADO.NET in order to practice about xml-object-relational databases: design, SQL, APIs and some brief examples about JSP and SPARQL.

### *Teaching Programme*

#### **I. Introduction to Databases. The ANSI/X3/SPARC Standard DBMS Architecture.**

#### **II. Conceptual Modelling and Semantic Data Models.**

- II.1. Introduction to Conceptual Modelling Abstractions and Data Models.
- II.2. The 'Entity-Relationship' data model: Definitions and Concepts.

#### **III. Relational Databases.**

- III.1. Introduction to classical data models in DBMS: Relational, Network (Codasyl) and Hierarchical Models.
- III.2. Overview of the Systems and Languages in classical database models: Relational, Codasyl and Hierarchical Databases.
- III.3. The Relational Data Model: Concepts and Definitions.
- III.4. Formal Languages of the Relational Model: Relational Algebra and Relational Calculus.

#### **IV. Relational Database Design.**

- IV.1. Introduction to the relational database design.
- IV.2. Introduction to the theory of relational databases.
- IV.3. Functional Dependencies.
- IV.4. Normal Forms for Relational Schemata.

#### **V. SQL, the Standard Object-Relational and XML Language**

- V.1. Introduction to the Relational Query Languages.
- V.2. SQL: Data Definition Language, DDL. BNF Notation.
- V.3. Pre-compiling and execution of a SQL Module (embedded in a host programming language).
- V.4. SQL: Data Manipulation Language, DML. BNF Notation.
- V.5. SQL: View Definition Sentences.
- V.6. SQL: Query Expressions. SQL Grammar. BNF Notation.
- V.7. SQL:92 and SQL:99 Standardizations.
- V.8. SQL:2003 Object-Relational and SQL/XML Data Models. Current Database Technology.
- V.9. SQL/XML:2006 and XML. Examples and Exercises.



## VI. Interoperability of Heterogeneous and Multiple Distributed Databases.

- VI.1. Concept of Interoperability among Heterogeneous Information Systems.
- VI.2. Interoperability of DBMS: RDA, ODBC, JDBC and C/S (net, open, star).
- VI.3. Multiple and Heterogeneous Database Systems: Autonomy, Heterogeneity & Distribution.
- VI.4. Homogeneous and Distributed Relational Database Organizations.
- VI.5. Web Databases Organization. The Mediator-Wrapper Architecture.
- VI.6. The deep Web with intensive Databases. The Semantic Web and SPARQL (2008).

## VII. Distributed DBMS Architecture.

- VII.1. Transparencies in Distributed DBMS.
- VII.2. DBMS Standardization. Architectural Models for Distributed DBMS.
- VII.3. Client/Server Architecture (two tiers) and Web (three or more tiers).
- VII.4. Semi-structured Data and Information Systems for Querying Heterogeneous Web data sources.

## VIII. Distributed Database Design.

- VIII.1. Alternatives of Design Strategies and Distribution Data Aspects.
- VIII.2. Fragmentation Algorithms. Data localization programmes.

## IX. Semantic Data Control.

- IX.1. View Management. Data Security.
- IX.2. Semantic Integrity Control.
- IX.3. Relevant objects for the semantics of ORDB and XML in SQL:2006.
- IX.4. Triggers and other Semantic Control Techniques. Exercises.

## X. Distributed Query Processing.

- X.1. Goals and Layers of Query Processing.
- X.2. Global Query Decomposition and Localization of Local Distributed Data.

## XI. Theory of Concurrency. Distributed Transaction Management and Concurrency Control.

- XI.1. Introduction to the Concurrency. The Interference Problem.
- XI.2. Definition, Properties and Types of Transactions.
- XI.3. Serializability Theory.
- XI.4. Concurrency Control, Techniques and Algorithms: *Locking, Timestamping, Optimistic and Pesimistic Approaches, Deadlock Management.*

## XII. Distributed DBMS Reliability.

- XII.1. The Concept of the Persistence, Measure of the Reliability.
- XII.2. Data Manager. Recovery and Persistence Techniques: *Log and Buffer Management, Chekpoint, Undo/Redo, etc.*
- XII.3. Local and Distributed Reliability Protocols (*Two-phase Commit Protocol*).

## Bibliography

### Text Books and documents:

1. **Sistemas de Bases de Datos. Conceptos, Técnicas y Lenguajes**, C. Costilla, S. Publicaciones E.T.S. Ingenieros de Telecomunicación-UPM, ISBN: 84-7402-271-1, 465 páginas, 1999
2. **Principles of Distributed Database Systems**, M. Tamer Özsu and P. Valduriez, Prentice-Hall, 2<sup>nd</sup> edition, ISBN: 0-13-607938-5, 666 pages, 1999.
3. **Complementary Documentation**, published at the Web page <http://sinbad.dit.upm.es>

**Coordinator and Professor:** Dra. Carmen Costilla Rodríguez

**Home page:** <http://sinbad.dit.upm.es>

**e-mail:** [costilla@dit.upm.es](mailto:costilla@dit.upm.es), **Office:** C-216, **Laboratory:** B-205/1

**Time-table:** Thursday (18 - 20 h.) and Friday (16 - 18 h.)

**Tutorial:** Thursday (20 - 22 h.) and Friday (18 - 20 h.)