



Bases de Datos Tema 2: Modelos de datos

Dpto. de Ingeniería Informática

Contexto

- Asignatura "Bases de datos", Grado en Ingeniería Informática UCA
- Competencias
 - C12 Conocimiento y aplicación de las características, funcionalidades y estructuras de las bases de datos, que permitan su adecuado uso, y el diseño y el análisis e implementación de aplicaciones basadas en ellos
 - C13 Conocimiento y aplicación de las herramientas necesarias para el almacenamiento, procesamiento y acceso a los Sistemas de información, incluidos los basados en web
- Resultados del aprendizaje
 - R2 Saber utilizar los diferentes modelos de datos, sus componentes y la importancia en el diseño de una base de datos
 - R3 Capacidad para descubrir la necesidad de utilizar los sistemas de bases de datos y presentar las características diferenciadoras de los mismos
- Tiempo estimado: 2 horas

Contenidos

- Introducción
- Funciones de un SGBD
- Componentes del entorno de un SGBD
- Arquitectura de una BD
- Modelos de Datos
- Aspectos de un MD
- Referencias

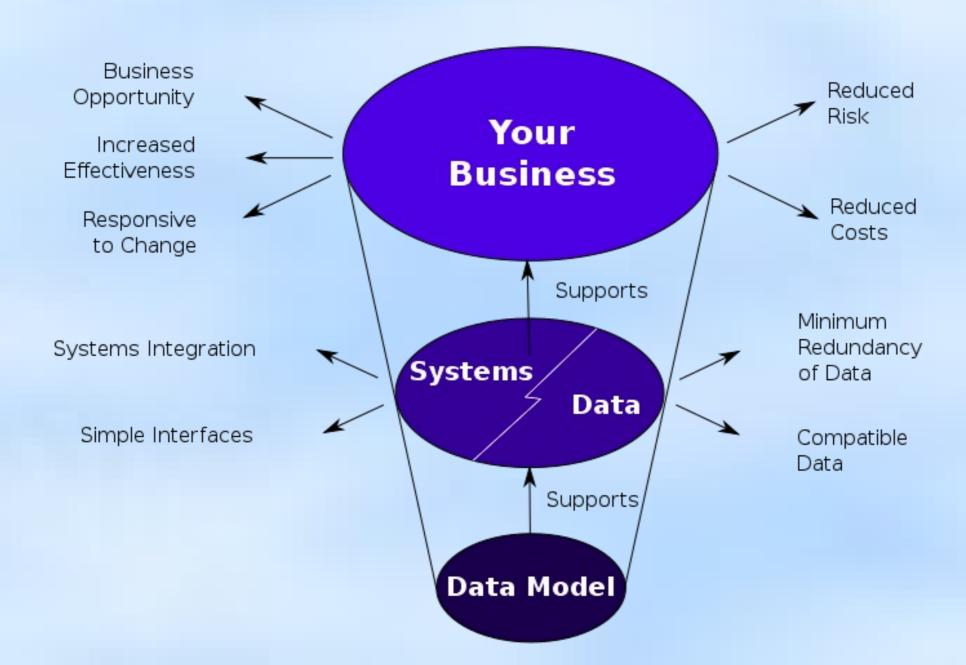
Introducción

- SGBD: Conjunto de programas que permiten a los usuarios crear y mantener una(s) BD
 - Hace de interfaz entre el usuario y la BD
- Facilita las tareas de:
 - Definición: los tipos de datos, y las restricciones existentes entre ellos
 - Construcción: proceso de introducción de los datos en la BD previamente definida
 - Manipulación: insertar, actualizar, eliminar y consultar (generación de informes) datos - CRUD

Sistemas Gestores de BD

- Los SGBD se construye a partir de un modelo de datos
- Modelo de Datos (MD): conjunto de conceptos y reglas que permiten estructurar los requisitos de datos de un sistema
 - Ese estudio lo realiza un humano que decide qué (y cómo) se incluye en la BD y qué se descarta
 - El MD se basa en estructuras de datos y operadores que ofrece el SGBD

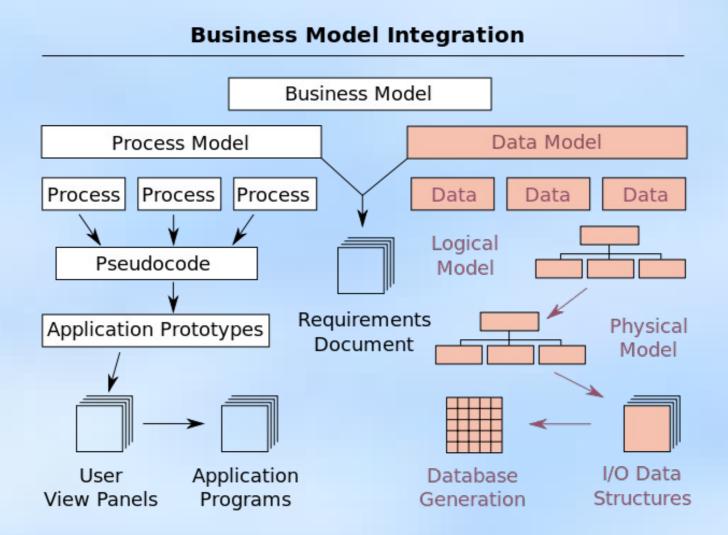
Beneficios del modelo de datos



- Independencia con respecto a los datos:
 - Capacidad de modificar un esquema de la BD sin necesidad de modificar el esquema del nivel inmediato superior
 - Se concreta en:
 - Independencia lógica: capacidad de modificar el esquema conceptual sin alterar los esquemas externos ni los programas
 - Independencia física: capacidad de modificar el esquema interno sin tener que alterar el esquema conceptual/lógico o los externos (ni programas)

Contexto de modelado de datos

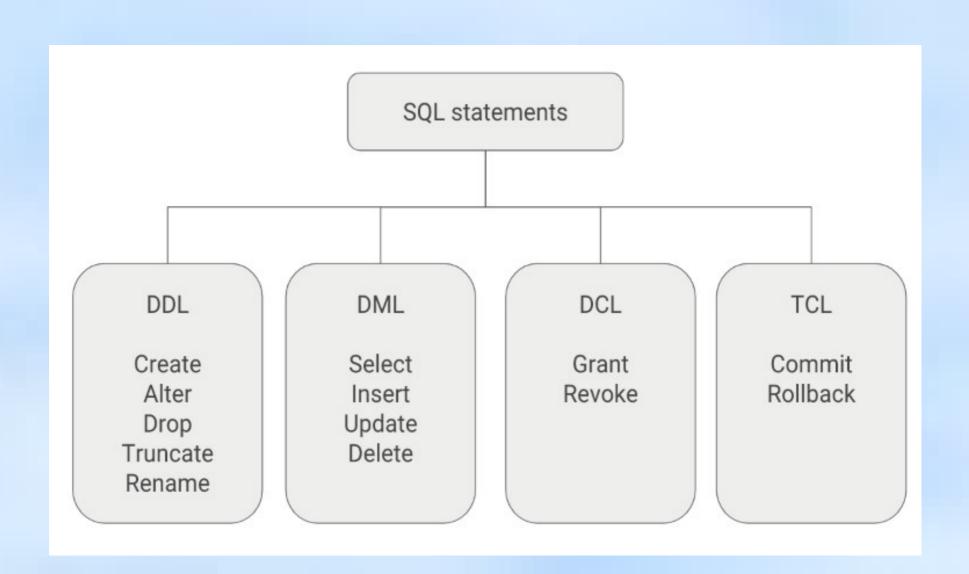
https://en.wikipedia.org/wiki/File:Data_modeling_context.svg



Sistemas Gestores de BD

- Objetivos de un SGBD:
 - No redundancia (no deseada) de los datos
 - Se comparten los datos por aplicaciones y/o usuarios
 - Manipulación de datos eficiente
 - Eficacia en acceso a los datos (según usuario)
 - Administración "centralizada" de los datos
 - Coherencia de los datos: mediante reglas de integridad
 - Compartición de los datos entre usuarios y aplicaciones (con acceso concurrente)
 - Seguridad de los datos

- El SGBD dar servicio a usuarios y aplicaciones
- Las operaciones típicas de un SGBD pueden afectar a:
 - La totalidad de los datos de una BD: suelen bloquearla
 - Creación
 - Reestructuración
 - Consulta a la totalidad
 - Sólo a ciertos datos: pueden ejecutarse en paralelo
 - Actualización: inserción, borrado y modificación
 - Consulta selectiva
- Se organizan en cuatro funciones:
 - Descripción, Manipulación y Control (y Transacciones)



- Función de Descripción o Definición:
 - Debe permitir al diseñador especificar:
 - Los elementos de datos de la BD
 - Estructura de los datos
 - Las relaciones que existen entre los datos
 - Las reglas de integridad semántica
 - Las característica de tipo físico
 - Las vistas lógicas de los usuarios
 - Se realiza con un Lenguaje de Definición de Datos
 (DDL) del SGBD que permite definir los tres niveles de la arquitectura y se almacena en el Dicc. de Datos
 - El SGBD se ocupa de la correspondencia entre estos niveles

- Función de Descripción o Definición:
 - Ejemplo:

```
45
    -- CREATE
26
   □CREATE TABLE Person(
28
        PersonID INT IDENTITY (1,1) CONSTRAINT PK PersonID PRIMARY KEY,
         FirstName NVARCHAR(20),
29
         LastName NVARCHAR(25)
30
31
32
33
    -- ALTER
   ALTER TABLE Person
35
    ADD BirthDate DATETIME
36
37
     -- DROP
     DROP TABLE Person
38
39
```

- Función de Manipulación:
 - La consulta a la BD puede ser:
 - Totalidad de los datos: se recuperan todos los datos o los de un determinado tipo (limita concurrencia)
 - Consulta selectiva: se recuperan los datos que cumplan un criterio
 - Actualización: inserción, borrado o modificación
 - Esta función se lleva a cabo con el Lenguaje de Manipulación de Datos (DML):
 - Lenguaje incorporado en el SGBD: SQL
 - Lenguajes huéspedes sobre un lenguaje anfitrión: SQL embebido en C/C++, Python, Javascript, PHP, ...

- Función de Manipulación:
 - Ejemplos:

//Program Segment E1:

```
1 DELETE FROM customers
2 WHERE store_state = 'MH'
3 AND customer_id = '1001';

Data Output Messages Notifications Explain

DELETE 1

Query returned successfully in 48 msec.
```

```
0) loop = 1;
    while (loop) {
      prompt("Enter a Social Security Number: ", ssn);
 2)
 3)
     EXEC SOL
 4)
        select Fname, Minit, Lname, Address, Salary
 5)
        into :fname, :minit, :lname, :address, :salary
 6)
        from EMPLOYEE where Ssn = :ssn ;
 7)
      if (SQLCODE == 0) printf(fname, minit, lname, address, salary)
 8)
        else printf("Social Security Number does not exist: ", ssn) ;
 9)
      prompt("More Social Security Numbers (enter 1 for Yes, 0 for No): ", loop);
10)
```

- Función de Control:
 - Interfaces usuarios BD
 - Mediante Lenguaje de Control de Datos (DCL)
 - Procedimientos para la administración de la BD, como:
 - Servicio:
 - Creación y mantenimiento de ficheros
 - Obtención de estadísticas de utilización de la BD
 - Gestión de usuarios
 - Seguridad física:
 - Copias de seguridad
 - Arranque y parada de la BD en casos de fallos
 - Protección contra accesos no autorizados

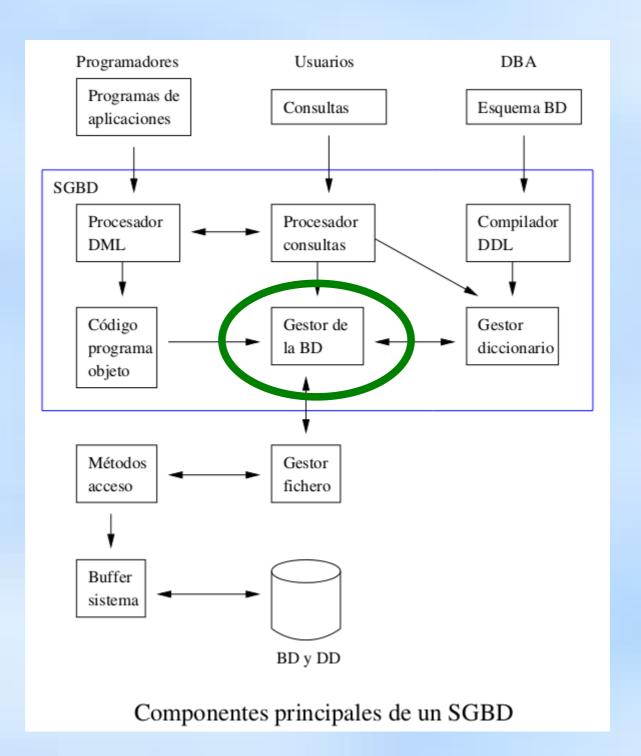
- Función de Control:
 - Ejemplos:

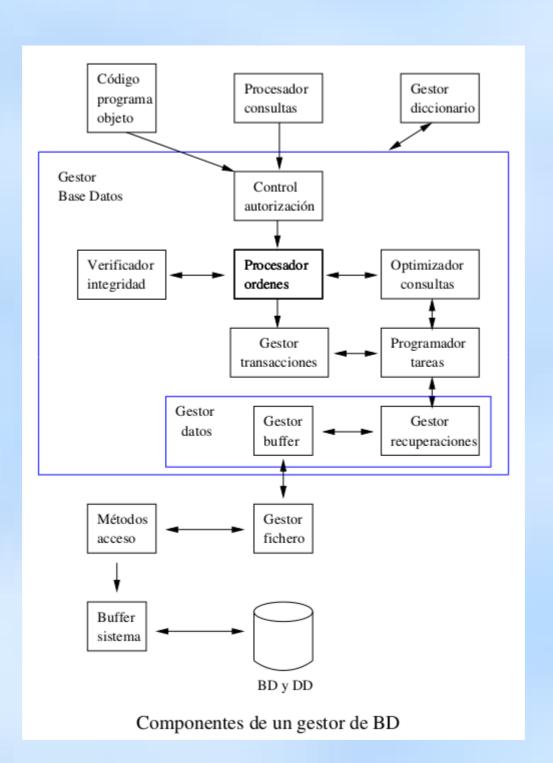
```
40
     -- GRANT
41
    GRANT SELECT, INSERT, UPDATE, DELETE ON Employees TO almir
42
43
44
     -- RFVOKE
     REVOKE INSERT ON Employees TO almir
45
46
     -- DFNY
47
    DENY UPDATE ON Employees TO almir
48
49
```

- Función de Transacciones:
 - Especificar transacciones
 - Mediante el Lenguaje de Control Transaccional (TCL)
 - Instrucción compleja ejecutada atómicamente
 - Procedimientos para la gestión, como:
 - COMMIT: vuelca a disco los datos en memoria
 - ROLLBACK: deshace las modificaciones desde el último COMMIT
 - Si se indica un SAVEPOINT se deshace hasta ese SAVEPOINT
 - SAVEPOINT: da nombre a un estado de datos al que se puede volver haciendo un ROLLBACK ~ GOTO
 - Se hace guardando el estado en memoria principal (no en disco)

- Función de Transacciones:
 - Ejemplos:

```
-- Start a new transaction
 BEGIN TRANSACTION
 -- SQL Statements
  INSERT INTO Product VALUES(116, 'Headphone', 2000, 30)
  UPDATE Product SET Price = 450 WHERE Product id = 113
  -- Commit changes
 COMMIT TRANSACTION
Messages
(1 row affected)
```





- Clasificación:
 - Máquina:
 - Software
 - Hardware
 - Humana:
 - Procedimientos
 - Personas
 - Datos: lazo de unión entre máquina y personas
- Profundizamos en cada uno

Máquina:

- Software: herramientas de 4^a generación (4GL) específicas de dominio
 - El lenguaje SQL permite manipular datos relacionales fácilmente
 - Pero no es sencillo hacer un programa para calcular Fibonacci
 - Lenguaje NO procedimental (demasiado) fácil de entender
 - Se basa en álgebra y cálculo relacional (potente)
 - Además suelen incorporar: generadores de gráficas, asistentes para informes, formularios para aplicaciones ...
 - No se ven en este curso
- Hardware: los SGBD tienden a ser independientes del hardware y de los SO ...
 - Aunque los fabricantes suelen tener sus SSOO <u>certificados</u>

• Humana:

- Procedimientos: para realizar copias de seguridad, herramientas del SGBD para ayuda a los usuarios, etc.
 - Tienen que estar documentados
 - https://twitter.com/bartpl/status/1369580887202013184

- Roles:

- Administrador de Datos (DA): LOPD/GDPR, ... (full disclosure)
- Administrador de la BD (DBA)
- Diseñadores de la BD
- Programadores de aplicaciones
- Usuarios finales (*lusers*)
- En organizaciones pequeña los roles se agrupan ...
 - Hasta llegar al informático orquesta

- Humana:
 - Es un elemento clave
 - Según el Enterprise Data Security Survey del Independent Oracle Users Group (2013):
 - http://www.oracle.com/us/products/database/security/201
 3-ioug-data-security-survey-2075568.pdf

• El riesgo más grande es el factor humano (por delante

de hacks)

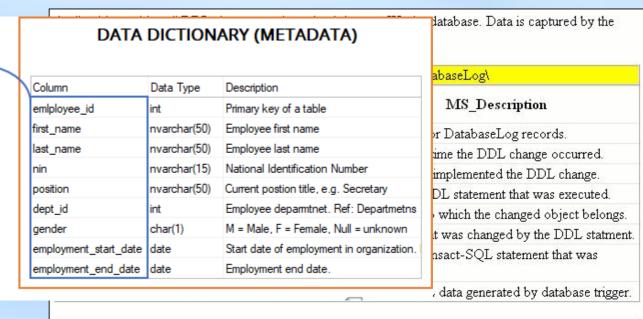


Datos:

- Ficheros de datos: contienen los datos en sí
- Diccionario de Datos (DD): es un repositorio de metadatos. Almacena información sobre los datos de nuestra BD:
 - Ficheros que almacenan los datos (rutas, formatos, etc)
 - Tablas
 - Tipos de datos
 - Valores que admiten
 - Relaciones entre datos
 - etc

Ejemplo de Diccionario de Datos:

DATA				
emlployee_id	first_name	last_name	nin	dept_id
44	Simon	Martinez	HH 45 09 73 D	1
45	Thomas	Goldstein	SA 75 35 42 B	2
46	Eugene	Comelsen	NE 22 63 82	2
47	Andrew	Petculescu	XY 29 87 61 A	1
48	Ruth	Stadick	MA 12 89 36 A	15
49	Barry	Scardelis	AT 20 73 18	2
50	Sidney	Hunter	HW 12 94 21 C	6
51	Jeffrey	Evans	LX 13 26 39 B	6
52	Doris	Berndt	YA 49 88 11 A	3
53	Diane	Eaton	BE 08 74 68 A	1



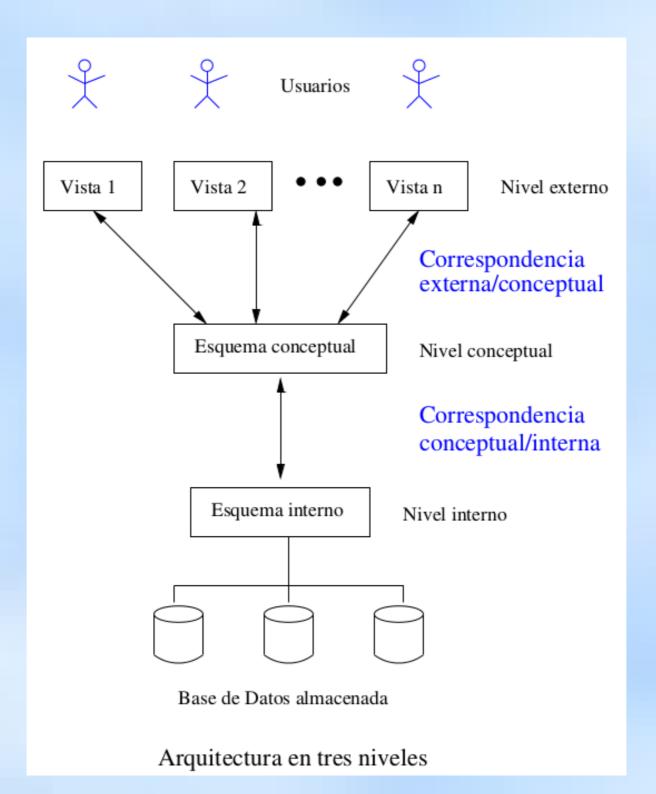
index_name	index_description		index_keys
PK_DatabaseLog_DatabaseLogID	_DatabaseLogID nonclustered, unique, primary key located on PRIMARY		DatabaseLogII
	CONSTRAINTS		
Туре	Name	Status	Keys
PRIMARY KEY (non-clustered)	PK_DatabaseLog_DatabaseLogID	(n/a)	DatabaseLogID
	FOREIGN KEYS		
No Foreign references to this table.			

Clasificación de los SGBD

- Existen clasificaciones según:
 - El MD en que está basado:
 - Relacional, Orientado a objetos, etc
 - Número de usuarios: monousuario vs multiusuario
 - Según el almacenamiento de los datos:
 - Centralizado vs distribuido
 - Propósito: general o específico (geográficos, ...)
 - Autoadaptables: SageDB
 - http://dsail.csail.mit.edu/index.php/projects/

- Las BD tienen las siguientes características:
 - Separación entre los programas y los datos: no hay datos "hardcodeados" ni dependientes de ficheros concretos
 - Ejemplo: mayoría de edad con 18 años
 - Múltiples vistas para las diversas necesidades de usuarios
 - Empleo de un catálogo para el almacenamiento de la descripción de la BD
- Los SGBD respetan la arquitectura propuesta por ANSI/PARC y distinguen tres niveles:
 - Externo, conceptual (a.k.a. global o lógico) e interno
 - La definición de la BD a cada uno de los niveles se denomina esquema.

- Niveles estandarizados:
 - Nivel interno o físico: describe la estructura física de almacenamiento y los caminos de accesos a la BD (único)
 - Nivel conceptual o lógico: define la estructura de toda la BD sin especificar las estructuras necesarias (único)
 - Nivel externo: vista de los usuarios (el esquema externo es el conjunto de esquema parciales de cada usuario/aplicación)
- Los SGBD deben asegurar la independencia entre estos niveles, para ello tenemos:
 - Correspondencia externa/conceptual
 - Correspondencia conceptual/interna



Nivel externo



Inspector para Iberoamérica

Nivel conceptual

DNI	Nombre	Nacionalidad
1111	Leonel	Argentina
2222	Cristiano	Portuguesa

	cod_c	DNI_culpable	Delito	Multa
	aaa	1111	Evasión	3.000.000€
	bbb	1111	False. Doc.	1.000.000€
	CCC	2222	False. Doc.	1.000.000€
	ddd	2222	Desfalco	4.000.000€

Nivel interno

Base de datos centralizada

Nivel externo



DNI	Nombre	Nacionalidad	cod_c	Delito	Multa
1111	Leonel	Argentina	aaa	Evasión	3.000.000€
1111	Leonel	Argentina	bbb	False. Doc.	1.000.000€

OJO, NO es un informe es acceso (bidireccional) a los datos

Nivel conceptual

DNI	Nombre	Nacionalidad
1111	Leonel	Argentina
2222	Crictiana	Portuguesa
	Challano	i ortuguesa

cod_c	DNI_culpable	Delito	Multa	
aaa	1111	Evasión	3.000.000€	
bbb	1111	False. Doc.	1.000.000€	
ccc	2222	False, Doc.	1.000.000€	
alalal	2222	Deefelee	4.000.0000	
uuu		Desialed	4.000.000€	

Nivel interno

Base de datos centralizada

Nivel externo



Nacionalidad	Núm_c	Sum_Multa
Argentina	2	4.000.000€
Portuguesa	2	5.000.000€



Nivel conceptual

DN	Nombre	Nacionalidad
1111	eonel	Argentina
2227	Cristiano	Portuguesa

cod_c	DNI_culpable	Delito	Multa
aaa	1111	Evasión	3.000.000€
bbb	1111	False. Doc.	1.000.000€
CCC	2222	False. Doc.	1.000.000€
ddd	2222	Desfalco	4.000.000€

Nivel interno

Base de datos centralizada



Nivel conceptual

DNI	Nombre	Nacionalidad
1111	Leonel	Argentina
2222	Cristiano	Portuguesa

cod_c	DNI_culpable	Delito	Multa
aaa	1111	Evasión	3.000.000€
bbb	1111	False. Doc.	1.000.000€
CCC	2222	False. Doc.	1.000.000€
ddd	2222	Desfalco	4.000.000€

Nivel interno

Base de datos centralizada

Copia de respaldo









(Distintas credenciales)

Nivel conceptual

DNI	Nombre	Nacionalidad
1111	Leonel	Argentina
2222	Cristiano	Portuguesa

cod_c	DNI_culpable	Delito	Multa
aaa	1111	Evasión	3.000.000€
bbb	1111	False. Doc.	1.000.000€
CCC	2222	False. Doc.	1.000.000€
ddd	2222	Desfalco	4.000.000€

Nivel interno

Base de datos Distribuida 1/2 Base de datos Distribuida 2/2



- Modelo de Datos (MD): mecanismo para ocultar detalles del almacenamiento y gestión de datos a los usuarios no administradores
 - Tres niveles: interno, conceptual y externo
- Los MD que permiten describir:
 - Estructura de la DB
 - Tipos de datos
 - Interrelaciones entre los datos
 - Restricciones de los datos

Nivel externo:

- Es la "vista" que cada usuario tiene de la BD
- Este nivel describe la parte de la BD que es relevante para cada usuario concreto
- Solo se incluyen las entidades, atributos y relaciones que son necesarias para dicho usuario
 - · Y se indica cómo calcularlos a partir del nivel conceptual
- Cada (tipo de) usuario tiene una representación de los datos personalizada

- Nivel conceptual:
 - Es la visión intermedia entre los otros niveles
 - Describe todos los datos almacenados en la BD y las relaciones existentes entre ellos
 - Ofrece una vista completa de los requisitos de datos que cubre el sistema
 - Es independiente del tipo de almacenamiento físico
- Es el que estudiaremos en la asignatura

Nivel interno:

- Indica la representación física de los datos
- Busca una implementación física óptima según los parámetros de rendimiento de uso de la BD
- Considera aspectos como:
 - Asignación de espacio para datos
 - Estructuras auxiliares de búsqueda (índices, ...)
 - Descripciones de los registros
 - Ubicación de los registros en fichero(s)
 - Compresión y cifrado de datos

Categorías de modelos de datos

- Los MD se pueden categorizar en:
 - De alto nivel (o conceptuales): usan constructores cercanos al usuario final de la DB
 - MD representacionales o de implementación: usan conceptos no muy diferentes del almacenamiento, pero medianamente fáciles de entender por usuarios
 - De bajo nivel (o físicos): cercanos a los detalles de implementación

Otros

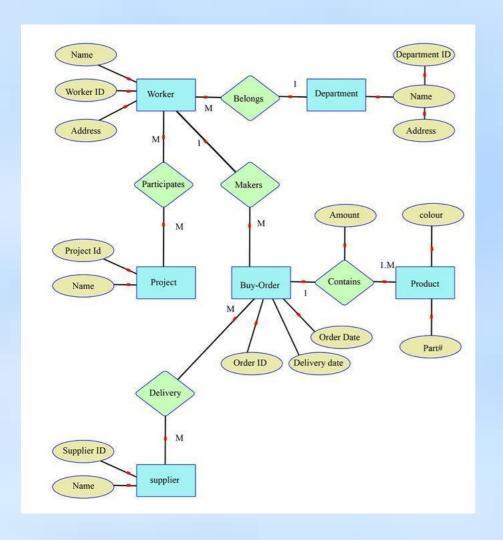
- Orientados a objetos: poco usado
- Autodescribibles: XML, key-value y otros NoSQL

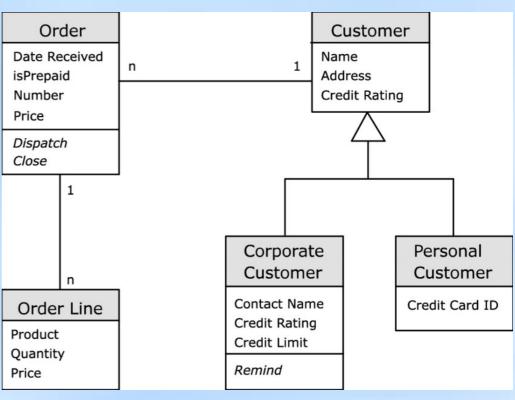
Modelos de datos conceptuales

- Los MD de alto nivel se basan en:
 - Entidades: que representan un concepto del mundo a modelar del que se almacenará alguna información en la base de datos
 - Atributos: describen características de interés de una entidad
 - Relaciones: indican las interdependencias entre dos o más entidades
- Se puede llamar también conceptuales
 - No son implementables
 - Son el paso previo al modelo representacional
- Los más usados son E/R y UML

Modelos de datos conceptuales

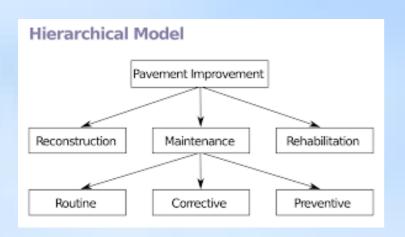
Ejemplos de MD de alto nivel: E/R y UML

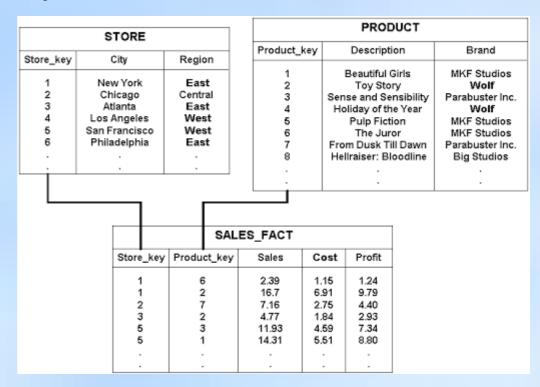




Modelos de datos representacionales

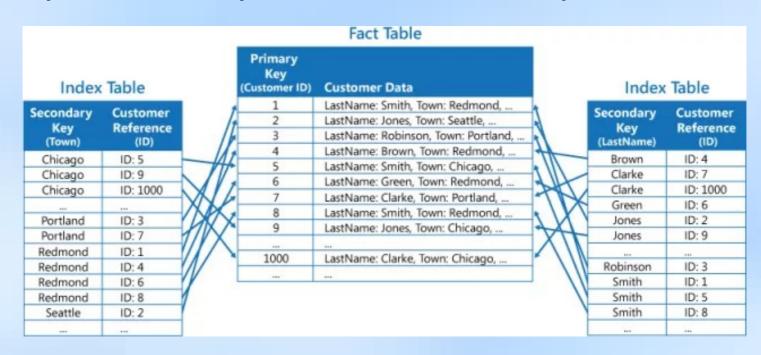
- Los MD representacionales:
 - Suelen ser los implementados en los SGBD
 - El más común es el MD relacional
 - Existen otros (en desuso):
 - jerárquico, en red, etc.





Modelos de datos de bajo nivel

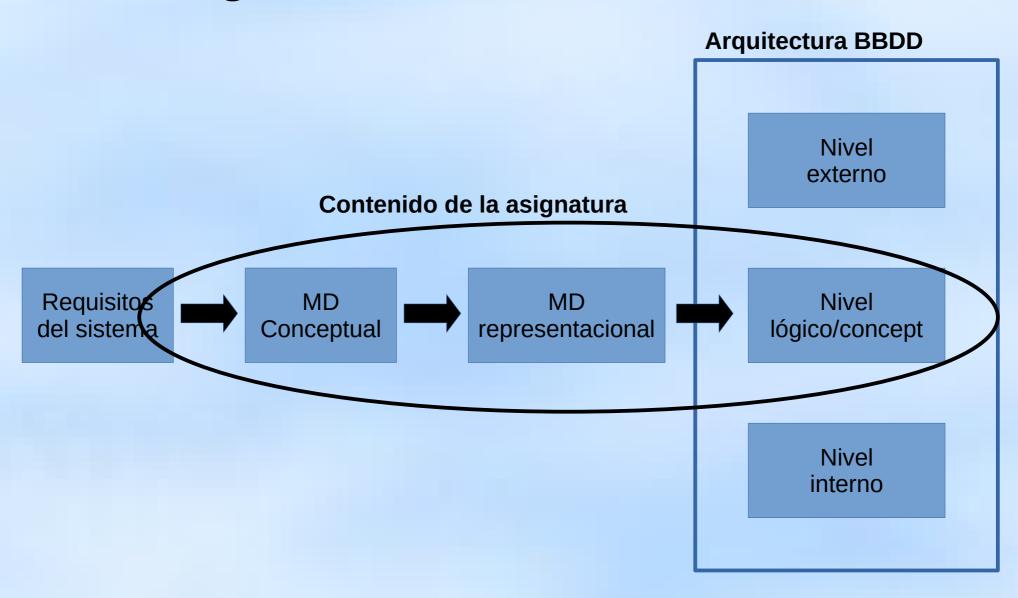
- Los MD de bajo nivel:
 - Se basan en estructuras de datos de bajo nivel
 - Ej: Árboles B+, estructuras con Hash, índices, etc
 - Concretan la implementación del nivel interno de la arquitectura a partir del modelo representacional



Categorías de modelos de datos

- En la asignatura (parte teórica) estudiaremos
 - El modelo conceptual E/R
 - Se construye a partir de requisitos del sistema
 - No ejecutable ni implementable directamente
 - El modelo representacional relacional
 - Se construye a partir del modelo conceptual E/R y otros requisitos
 - Se puede implementar en el nivel conceptual (o lógico) de la arquitectura estándar de una BBDD mediante el lenguaje DDL
- No se estudian los modelos para los niveles externos ni internos de la arquitectura
 - Quien esté interesado: asignatura "Administración de BBDD"
- No confundir los tres niveles estandarizados de arquitectura de una BBDD con los tres tipos de MD más usados

Categorías de modelos de datos



- Todo modelo de datos tienen dos aspectos:
 - Estáticos: descripción de la base de datos (a.k.a. esquema de la BD)
 - No debe cambiar demasiado en el tiempo
 - Si hay demasiados cambios en el esquema de una BD ver alternativas a los modelos "clásicos"
 - Dinámicos: los datos que en un momento se almacenan (instancias u ocurrencias). Deben cumplir el esquema
 - Pueden cambiar mucho en el tiempo

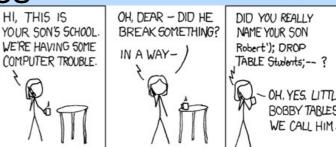
- Limitaciones de aspectos estáticos de un MD:
 - El MD permite definir y manipular elementos como objetos, asociaciones, propiedades, dominios, etc
 - No debe cambiar demasiado en el tiempo
 - Elementos que escapan del ámbito del MD usado:
 - Restricciones inherentes: por falta de expresividad en el MD (el humano sabe qué quiere poner pero el sistema no tiene constructores para ello)
 - Ej: el dato puede almacenar números enteros múltiplos de 7
 - Restricciones semánticas o de integridad: porque sean restricciones propias de los datos tratados (qué significan los datos en sí, son problemas humanos)

- Restricciones semánticas y de integridad:
 - Las semánticas atienden al significado de los datos en el mundo real
 - Las de integridad atienden a la corrección de los datos y su consistencia con el mundo real
 - Suelen ir ligadas: una correcta descripción semántica permite que comprobar la integridad de los datos (que puede hacer el SGBD)
- Ejemplo: toda ciudad pertenece a una (y sólo una) provincia

- Limitaciones de aspectos dinámicos de un MD:
 - Los datos que admite una BD están restringidos por su MD
 - En cada momento del tiempo pueden cambiar
 - Ojo: cuidado con la concurrencia (ejemplo: last_intert_id)
 - Evolucionan usando operadores que proporciona el DM para cada tipo de elemento
 - No son iguales los operadores para objetos que para atributos
 - Cada operación se define por un par (localización, acción)

- Ejemplo de Esquema de la BD (estático)
 - Alumno: DNI, nombre, apellidos, curso
 - Asignatura: código, nombre, créditos
- Ejemplo de base de datos (dinámico):
 - Alumno (12345678, Pedro, Pérez, 1)
 - Alumno (12345679, Ana, López, 2)
 - Asignatura (17101, Álgebra, 6)
 - Asignatura (17103, Bases de datos, 6)
- Alumno y Asignatura son constructores del esquema

- Ventajas de integrar las restricciones en la definición de datos:
 - Gestión centralizada:
 - La realiza el diseñador de la base de datos
 - Si se incluyen en el esquema, el SGBD se encargará de velar porque los usuarios las cumplan
- Alternativa: delegar en programas
 - Que cada software que use la BD almacene y compruebe las restricciones
 - https://xkcd.com/327/





Una reflexión:

- Hasta ahora en la carrera hemos usado lenguajes de programación que alguien definió por nosotros
 - Con estructuras de control "clásicas": if, while, etc
 - Con tipos de datos simples limitados
 - Pero esos tipos de datos se podían ampliar (por ejemplo, creando el tipo de dato "número complejo")
 - Pero si guardo los número complejos en un fichero ¿puedo poner restricciones al uso que hacen de él varios programas?

Cuestiones

- En los modelos de datos:
 - ¿Suele haber capacidad para definir tipos en tiempo de ejecución?
 - ¿Se pueden indicar restricciones de acceso?
 - ¿Se puede exigir rendimiento?

_

Referencias

- Apuntes de bases de datos de la profesora Esther Gadeschi 2003/4
 - Apuntes de bases de datos del profesor Carlos Rioja 2006/7
- Libro "Fundamentals of Database Systems"
 Elmasri y Navathe (3a Ed.)

Gracias por la atención ¿Preguntas?