

# Análisis y Diseño de Bases de Datos

---

## Introducción

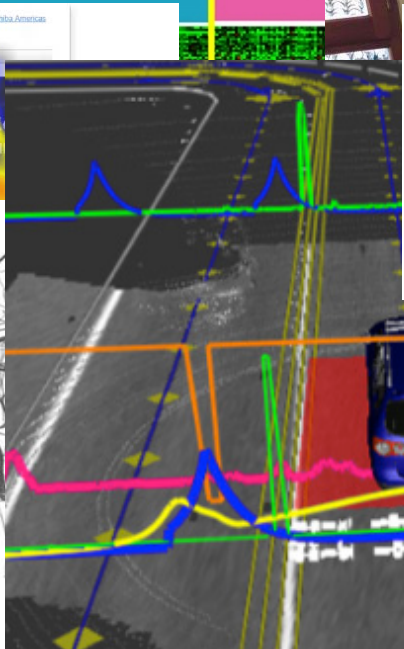
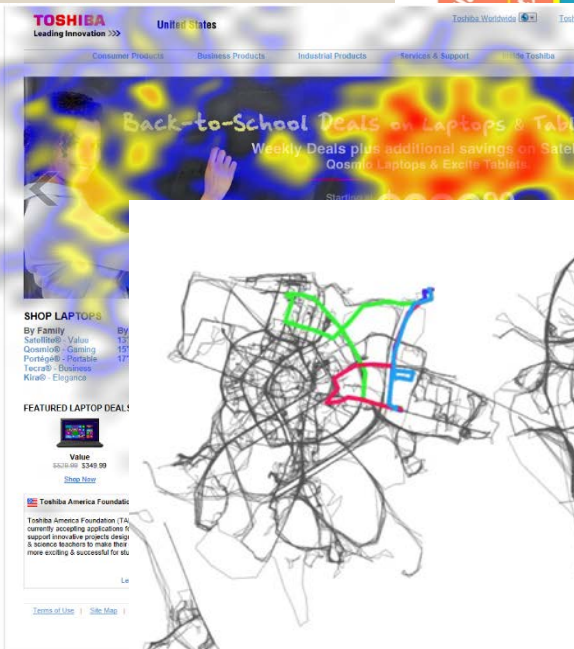
# Datos

- Representación o codificación de algún tipo de información o conocimiento
  - Datos: valores que se guardan y analizan
  - Información: datos analizados que sirven para tomar decisiones (interpretación)
  - Conocimiento: compresión de la información en base a experiencia sobre una materia
- Dato: registro de alguna acción
  - Cualquier cosa puede registrarse
  - Distintos tipos de datos



Rank	Site	Both			Site
		Incidence rate (1/10 <sup>5</sup> )	%	ASIRC* (1/10 <sup>5</sup> )	
1	Lung (C33-34)	53.57	18.74	25.34	Lung (C33-34)
2	Stomach (C16)	36.21	12.67	17.85	Stomach (C16)
3	Colon, rectum	29.44	10.30	14.21	Liver (C22)

Rank	Site	Both			Site
		Incidence rate (1/10 <sup>5</sup> )	%	ASIRC* (1/10 <sup>5</sup> )	
1	Lung (C33-34)	53.57	18.74	25.34	Lung (C33-34)
2	Stomach (C16)	36.21	12.67	17.85	Stomach (C16)
3	Colon, rectum	29.44	10.30	14.21	Liver (C22)



# Preguntas clave

- Qué queremos que haga la BD con estos datos en términos de guardarlos y analizarlos?
  - solo nos interesan los datos *persistentes* en memoria secundaria
  - guardarlos puede ser relativamente estándar (*ficheros binarios*), pero analizarlos es tan variado como los propios datos que tratemos
- Se puede construir un sistema de base de datos que pueda hacer esto para cualquier tipo de dato que nos encontremos?
  - no existe un único sistema que pueda gestionar cualquier cosa

# Por dónde empezamos?

- Por los datos **más fáciles**
  - hay muchos de estos? **sí, pero cada vez más de los complejos** — **revolución big data**
  - esto valdrá de algo cuando queramos tratar los complejos? sí, pero con limitaciones



**Automatricula**  
2005-2006  
Créditos 31.5 | Tiempo Restante 0h 07

Inicio | Datos Personales | **Asignaturas** | Orden de Desmatriculación | Actividades | Cierre Matricula | [Cerrar sesión](#)

Asignaturas

Tipo de bloque: Troncales y Universidad  
Núm. de bloque: PRIMERO

Cod.	Asignatura	Qu
✓ 5544	ESTRUCTURA Y TECNOLOGÍA DE COMPUTADORES	T
✓ 5545	PROGRAMACIÓN	T
✓ 5548	MATEMÁTICA DISCRETA Y ÁLGEBRA	T
✓ 5549	ANÁLISIS MATEMÁTICO	A
✓ 5549	INGLÉS TÉCNICO	A
✓ 5551	COMPUTACIÓN NUMÉRICA	B
✓ 5552	ESTADÍSTICA	B
✓ 5553	FUNDAMENTOS FÍSICOS DE LA INFORMÁTICA	A
? 5554	AMPLIACIÓN DE TECNOLOGÍA DE COMPUTADORES	B

Matriculados ✓ Aprobados ? Pendiente de acta

**Table 7** Top 10 cancer incidence rates in cancer registration areas in 2009

Rank	Both			Male			Female		
	Site	Incidence rate (1/10 <sup>5</sup> )	% ASIRC <sup>*</sup> (1/10 <sup>5</sup> )	Site	Incidence rate (1/10 <sup>5</sup> )	% ASIRC <sup>*</sup> (1/10 <sup>5</sup> )	Site	Incidence rate (1/10 <sup>5</sup> )	% ASIRC <sup>*</sup> (1/10 <sup>5</sup> )
1	Lung (C33-34)	53.57	18.74	25.34	Lung (C33-34)	70.40	Breast (C50)	42.55	16.81
2	Stomach (C16)	36.21	12.67	17.85	Stomach (C16)	49.61	Lung (C33-34)	36.34	14.36
3	Colon, rectum (C18-21)	29.44	10.30	14.21	Liver (C22)	41.99	Colon, rectum (C18-21)	26.42	10.44
4	Liver (C22)	28.71	10.04	14.78	Colon, rectum (C18-21)	32.38	Stomach (C16)	22.50	8.89
5	Esophagus (C15)	22.14	7.74	10.88	Esophagus (C15)	30.44	Liver (C22)	15.11	5.97
6	Breast (C50)	21.21	7.42	11.64	Prostate (C61)	9.92	Esophagus (C15)	13.64	5.39
7	Pancreas (C25)	7.28	2.55	3.35	Bladder (C67)	9.78	Cervix (C53)	12.96	5.12
8	Lymphoma (C81-85, 88, 90, 96)	6.68	2.34	3.75	Pancreas (C25)	8.24	Thyroid (C73)	10.09	3.99
9	Bladder (C67)	6.61	2.31	3.03	Lymphoma (C81-85, 88, 90, 96)	7.71	Uterus (C54-55)	8.77	3.46
10	Thyroid (C73)	6.56	2.29	4.21	Kidney (C64-66, 68)	7.07	Ovary (C56)	7.95	3.14
Top 10		218.40	76.39	109.05	Top 10		267.55	84.14	135.81
Top 10		196.32	77.57	99.01	Top 10		196.32	77.57	99.01

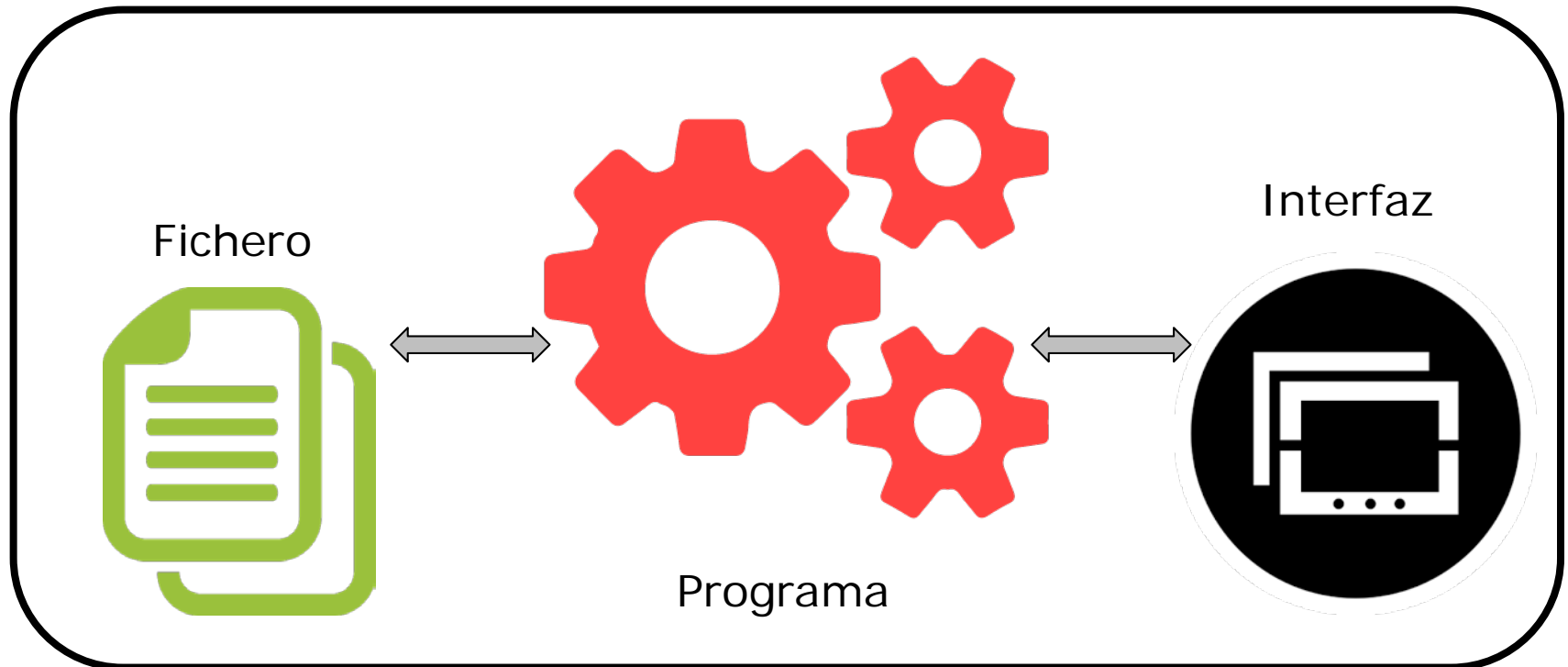
<sup>\*</sup>Age-standardized incidence rate (China population)

...

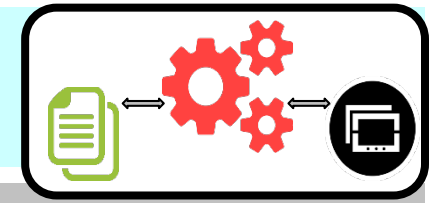
- Fáciles = tabla = *estructurados*
    - guardar: estructuras de datos orientadas a registros con campos, etc.
    - analizar: recuperar los datos, incluyendo condiciones lógicas, ordenaciones, agrupaciones, operaciones aritméticas
- ⇒ sistema de base de datos general que pueda tratar (*casi*) cualquier tipo de datos estructurado, acompañado de un lenguaje de consulta

# Sin BDs ...

- fichero con los datos; programa con acceso y lógica sobre los datos; interfaz de usuario



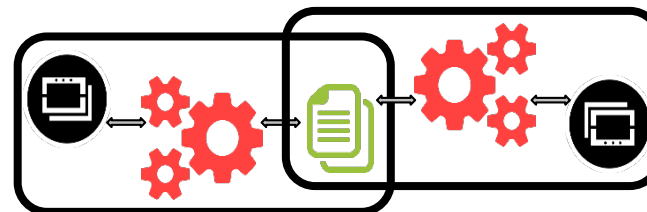
...



## ■ Problemas

- tipo de fichero: secuencial, tabla? eficiencia de las búsquedas?
- programa con acceso a fichero: abrir/cerrar, acceso a bloques; acceso a índices si cambio tipo de fichero, tengo que reprogramar
- soporta accesos concurrentes?
- como implemento la seguridad/privacidad/etc?
- qué ocurre con los datos erróneos?
- y cuando hay un fichero y varios programas?

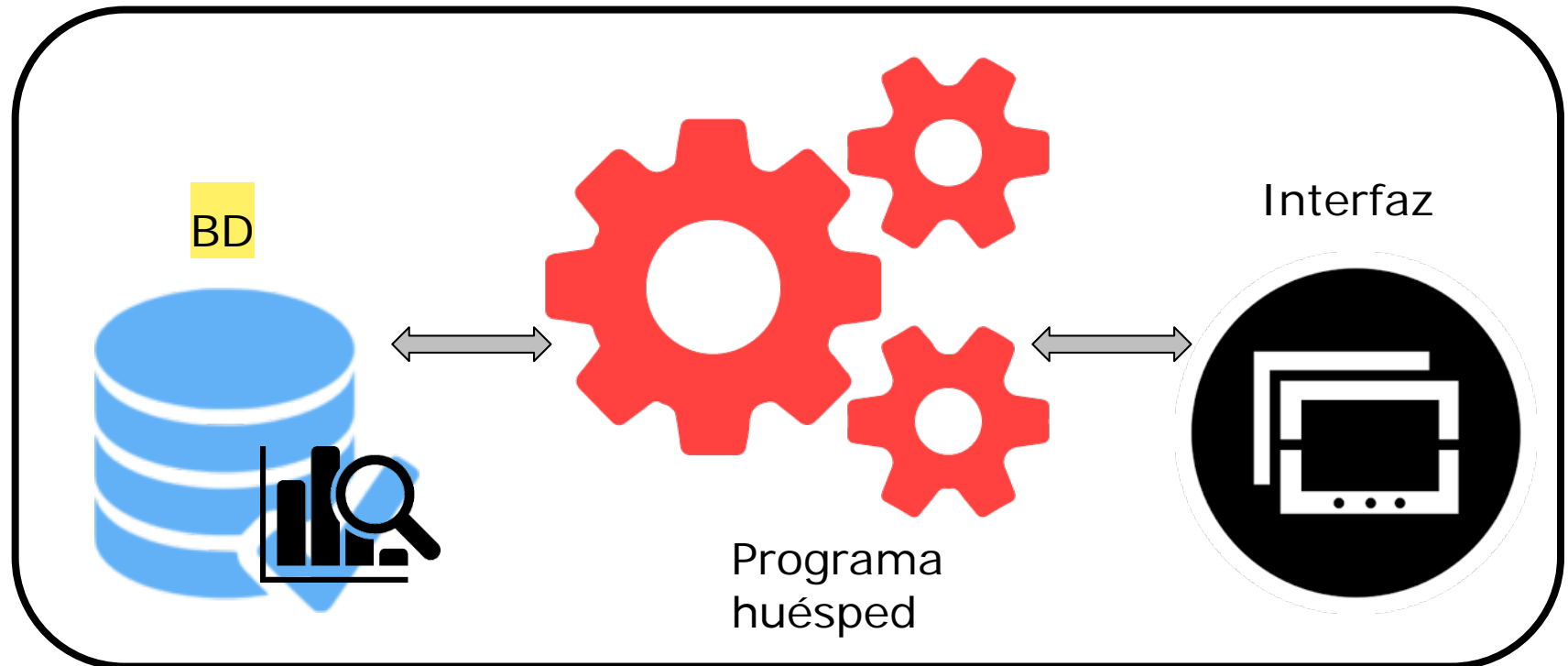
Rank	Male			Female						
	Site	Incidence rate (1/10 <sup>5</sup> )	% ASIRC (1/10 <sup>5</sup> )	Site	Incidence rate (1/10 <sup>5</sup> )	% ASIRC (1/10 <sup>5</sup> )				
1	Lung (C33-34)	53.57	18.74	25.34	Lung (C33-34)	34.75	Breast (C50)	42.55	18.81	
2	Stomach (C16)	38.21	12.87	17.85	Stomach (C16)	49.61	15.60	Lung (C33-34)	38.34	14.26
3	Colon, rectum (C18-21)	28.41	10.30	14.21	Liver (C22)	41.99	13.21	Colon, rectum (C18-21)	28.42	10.41
4	Liver (C22)	28.71	10.04	14.78	Colon, rectum (C18-21)	32.38	10.16	Stomach (C16)	22.50	8.86
5	Esophagus (C15)	22.14	7.74	10.88	Esophagus (C15)	30.44	9.87	Liver (C22)	15.11	5.97
6	Breast (C50)	21.21	7.42	11.64	Prostate (C61)	9.92	3.12	Esophagus (C15)	13.64	5.39
7	Pancreas (C25)	7.29	2.55	3.35	Bladder (C67)	9.78	3.05	Cervix (C23)	12.56	5.12
8	Lymphoma (C81-85, 88, 90, 98)	6.88	2.51	3.75	Pancreas (C25)	8.24	2.59	Thyroid (C73)	10.09	3.99
9	Bladder (C67)	6.61	2.31	3.03	Lymphoma (C81-85, 88, 90, 98)	7.71	2.42	Uterus (C54-55)	8.77	3.48
10	Thyroid (C73)	6.58	2.29	4.21	Kidney (C64-65, 68)	7.07	2.22	Ovary (C56)	7.55	3.14
Top 10		218.60	76.39	100.00	Top 10		267.55	84.14	130.81	
Top 10		218.60	76.39	100.00	Top 10		267.55	84.14	130.81	





# Con BDs ...

- No es un entorno general de programación / computación



# Base de datos, BD

- BD: colección organizada de datos, que
  - modela aspectos relevantes de la realidad, ej.: clientes, llamadas, tarifas
  - da soporte a procesos de información: ej.: 10% clientes con menor ratio fijo/móvil
- SGBD: sistema gestor de BD
  - sistema software que gestiona los datos
- Gestión: almacenamiento, extracción, modificación, borrado, búsqueda, seguridad, integridad, compartición, ...

# Nivel lógico

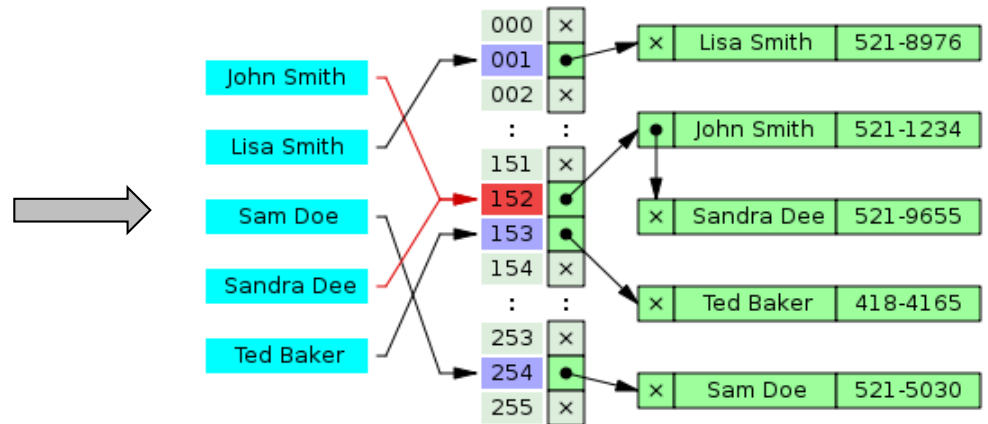
PATIENTS						
YPL-320	SMITH	m	38	176	1	124
GLI-532	JOHNSON	m	43	163	0	109
PNI-258	WILLIAMS	f	38	131	0	122
MIJ-579	JONES	f	40	133	0	117
XLK-030	BROWN	f	49	119	0	122
TFP-518	DAVIS	f	46	142	0	121
LPD-746	MILLER	f	33	142	1	130
ATA-945	WILSON	m	40	180	0	115

definimos, representamos y manipulamos datos como si fueran tablas —visión 'usuario'— aunque internamente/físicamente la BD los gestiona de otra forma

- la *traducción* al esquema interno la hace la BD → ventajas !!
- comprensible por agrupación de columnas
- estructura de tabla = entidades, propiedades y relaciones
- la línea 2 es especial = esquema de los datos
- problemas para almacenar otros datos en esa misma tabla

# Nivel físico

PATIENTS						
id	name	sex	age	wgt	smoke	sys
YPL-320	SMITH	m	38	176	1	124
GLI-532	JOHNSON	m	43	163	0	109
PNI-258	WILLIAMS	f	38	131	0	122
MIJ-579	JONES	f	40	133	0	117
XLK-030	BROWN	f	49	119	0	122
TFP-518	DAVIS	f	46	142	0	121
LPD-746	MILLER	f	33	142	1	130
ATA-945	WILSON	m	40	180	0	115



- estructuras de datos en memoria secundaria: bloques, índices, etc; métodos de acceso y coste asociado
- independencia lógico-física: automatizado por BD; se puede cambiar el diseño físico y mantener nuestros 'programas'

## Usuario (N. Lógico)

1. abrir tabla
2. seleccionar sys>120 & smoke
3. recuperar solo id,name
4. adjuntar conteo y promedio(sys)

## BD (N. físico)

1. abrir archivo
2. abrir índice, recorrer y comprobar registros sys, smoke
3. recuperar bloques y registros id, name
4. hacer cálculos y formatear salida

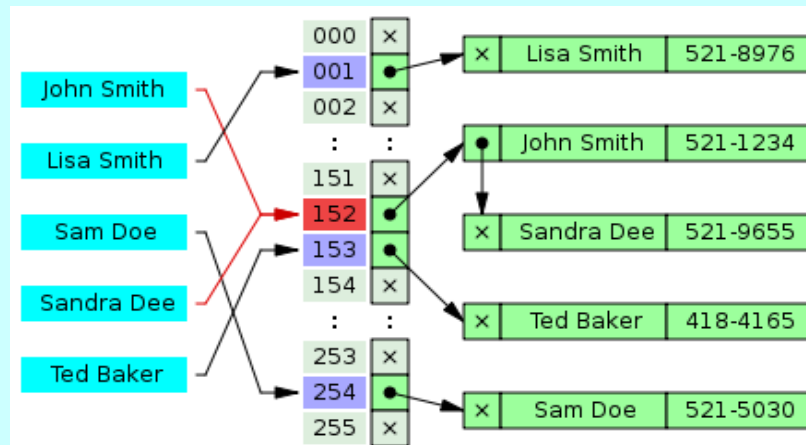
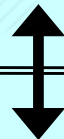
# INDEPENDENCIA LÓGICO-FÍSICA



Essentials



PATIENTS						
id	name	sex	age	wgt	smoke	sys
YPL-320	SMITH	m	38	176	1	124
GLI-532	JOHNSON	m	43	163	0	109
PNI-258	WILLIAMS	f	38	131	0	122
MIJ-579	JONES	f	40	133	0	117
XLK-030	BROWN	f	49	119	0	122
TFP-518	DAVIS	f	46	142	0	121
LPD-746	MILLER	f	33	142	1	130
ATA-945	WILSON	m	40	180	0	115



- traducción automatizada por la BD
- se puede cambiar el diseño físico, manteniendo los programas de usuario

# Nivel almacenamiento



- almacenamiento físico para persistencia
- la base de datos hará las peticiones de acceso al sistema operativo

# Estructura vs Complejidad

- añadir diversas mediciones sys

sys
124, 129, 117, ...

grupo repetitivo: n° indeterminado de valores en una celda

—las BD relacionales no lo permiten por eficiencia

YPL-320	SMITH	...	124
YPL-320	SMITH	...	129
YPL-320	SMITH	...	117

repetición de datos e inconsistencia:

YPL-320	SMITH	...
YPL-320	SMITZ	...

- añadir doctor(es) responsable(s) —nombre, especialidad—

YPL-320	SMITH	...	DrX, Alerg
PNI-258	WILLIAMS	...	DrX, Alerg
TFP-518	DAVIS	...	DrX, Card

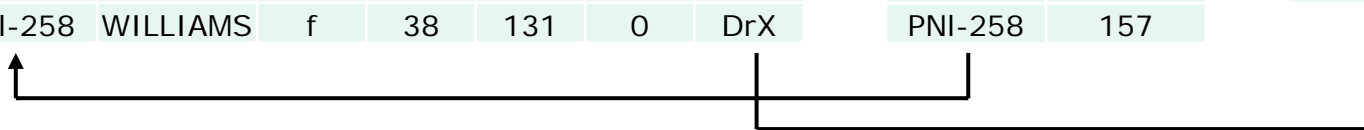
repeticiones, inconsistencia

- Utilizar varias tablas

PATIENTS						
id	name	sex	age	wgt	smoke	dr
YPL-320	SMITH	m	38	176	1	DrX
GLI-532	JOHNSON	m	43	163	0	DrY
PNI-258	WILLIAMS	f	38	131	0	DrX

SYS	
pat	val
YPL-320	124
YPL-320	129
PNI-258	157

DOCTOR	
name	special
DrX	Alerg
DrY	Card



# GRUPO REPETITIVO



Essentials

- Número indeterminado de valores en un atributo de una entidad

id	sys
GLI-532	124, 129, 117, ...



# INCONSISTENCIA



Essentials

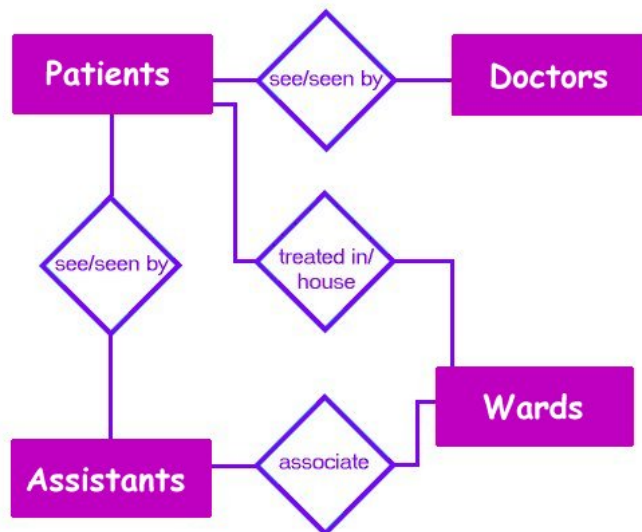
- Información repetida que da lugar a no coincidencia

id	name
YPL-320	SMITH
YPL-320	SMITZ

# Nivel conceptual

PATIENTS						
id	name	sex	age	wgt	smoke	sys
YPL-320	SMITH	m	38	176	1	124
GLI-532	JOHNSON	m	43	163	0	109
PNI-258	WILLIAMS	f	38	131	0	122
MIJ-579	JONES	f	40	133	0	117
XLK-030	BROWN	f	49	119	0	122
TFP-518	DAVIS	f	46	142	0	121
LPD-746	MILLER	f	33	142	1	130
ATA-945	WILSON	m	40	180	0	115

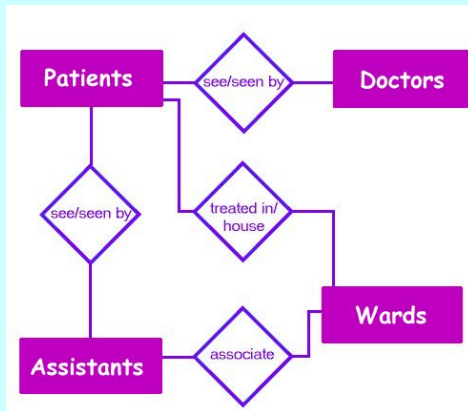
- no necesitamos ver las *instancias* de los datos
- modelado de la estructura de los datos
- visión global de alto nivel, basada en entidades, relaciones y propiedades
- o en clasificadores, asociaciones y atributos
- aclaración sobre los modelos/esquemas/diseños conceptuales y lógicos



# NIVELES / DISEÑOS



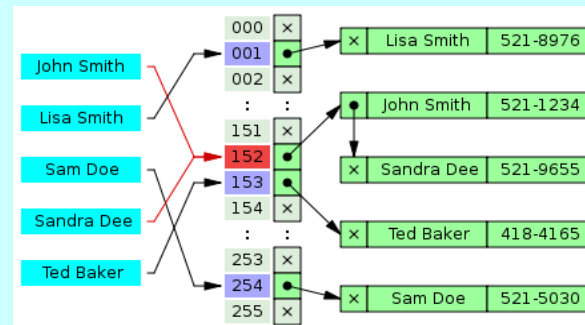
Essentials



## Conceptual

PATIENTS						
id	name	sex	age	wgt	smoke	sys
YPL-320	SMITH	m	38	176	1	124
GLI-532	JOHNSON	m	43	163	0	109
PNI-258	WILLIAMS	f	38	131	0	122
MIJ-579	JONES	f	40	133	0	117
XLK-030	BROWN	f	49	119	0	122
TFP-518	DAVIS	f	46	142	0	121
LPD-746	MILLER	f	33	142	1	130
ATA-945	WILSON	m	40	180	0	115

## Lógico



## Físico



## Almacenamiento

# Funcionalidades adicionales

1. **Integridad (corrección?) y seguridad de los datos**
  - el SGBD asegura las restricciones de integridad, en vez de los múltiples programas de acceso
  - controles de acceso a datos visibles para cada usuario
2. **Centralización de los datos**
  - minimizar la redundancia; evitar inconsistencias
3. **Acceso concurrente y recuperación**
4. **Reducción del tiempo de desarrollo y mantenimiento de aplicaciones**

# Control de datos incorrectos

## ■ Erróneo vs incorrecto

dni	nombre	fdn
147	ADELA	24/06/2000

en realidad es erróneo, Adela nació el 24/07/2000  
—pero es correcto (para la BD)

dni	nombre	fdn
147	ADELA	24/07/3000

ahora ya es incorrecto porque la BD puede comprobarlo (fecha > actual)  
—esto se denomina integridad

## ■ Integridad implícita

### 1. Dominio

sensor	temp (K)
S089	-52

### 2. Identificación

dni	nombre	fdn
147	ADELA	24/06/2000
147	TOMAS	02/12/1999

### 3. Referencial

PATIENTS		
id	name	...
YPL-320	SMITH	...
GLI-532	JOHNSON	...
PNI-258	WILLIAMS	...

SYS	
pat	val
YPL-320	124
YPL-320	129
BCT-330	157



# INTEGRIDAD



## Essentials

- Corrección de datos comprobable automáticamente por la BD

dni	nombre	fdn
147	Adela	24/07/3000

# Niveles de abstracción

- Esquema externo
  - *vista* (tabla) que combina datos para una presentación específica de usuario
  - se define sobre el conceptual
- Esquema conceptual
  - incluye todas las tablas/relaciones con información sobre entidades y relaciones
- Esquema interno
  - esquema físico con los detalles de almacenamiento

# Lenguajes de consulta

» formulación de consultas a través de un lenguaje

- Definición de esquemas (DDL), manipulación de datos (DML), control
- Formalización a través de cálculo/álgebra relacional (utilidad?)
- Optimización de la eficiencia
- SQL estándar
- Lenguaje anfitrión



# Arquitectura de un SGBD

- Bloques que desarrollan las funciones encargadas al SGBD
  - interacción con usuario/aplicación y almacenamiento secundario
  - secuencia operativa
  - interdependencia
  - configuración para mejorar eficiencia
  - trabajo del administrador de la BD

...

