

Análisis y Diseño de Bases de Datos

Introducción

Información y datos

- Información: el valor de los datos
- Muchas aplicaciones informáticas se dedican a gestionar datos
- Dificultad: por cantidad y por complejidad
 - millones de llamadas
 - clientes, tarifas, tickets
 - condiciones tarificación
 - fijo/móvil, horario, planes, roaming, datos, ...



Tipos de datos



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Daily mean flow values 2000													
2	Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	
3	1	4124	1474	2746	5192	4654	1012	4045	9641	6429	3130	3394	4776	
4	2	3154	1189	2910	2937	3711	4233	3312	18123	6471	2315	2778	20277	
5	3	7121	1105	22719	12916	1309	4051	3487	11312	10929	2056	2189	12560	
6	4	5693	1136	18710	16825	3099	6125	3334	9015	4017	14162	1240	8017	
7	5	4054	1109	19167	11010	2621	11158	3724	9017	5986	25108	1592	6680	
8	6	5083	1110	3907	11312	2361	8514	4302	7827	4829	11168	2825	6446	
9	7	2447	1064	6384	9674	3055	7315	4341	8515	5399	4942	1438	5395	
10	8	3098	1007	6337	7516	1879	7275	4746	17943	6103	5474	6110	4650	
11	9	1813	1041	6013	6126	1483	6612	17590	17129	7210	4397	11048	4137	
12	10	1009	1186	5116	9117	1370	5953	37967	11041	11819	8170	9199	4618	
13	11	1412	1141	6490	6190	1771	1712	29129	30008	17189	7995	2498	3405	
14	12	1180	807	3395	12718	1949	18105	29860	13919	19219	3717	6100	2252	
15	13	1523	719	3616	14872	1795	9710	17045	30618	11819	3101	1146	2181	
16	14	1015	1182	2643	10048	1416	7816	15343	20389	14136	5148	3008	1952	
17	15	954	1878	1645	7918	1277	4613	10960	11481	18718	7540	2647	1724	
18	16	2950	1181	4874	1180	1207	4611	8311	10114	17614	6114	2911	1406	
19	17	17098	1218	6275	4414	1117	6312	7068	10866	40791	4124	1190	1409	
20	18	16919	1817	6916	4018	1256	1860	5643	8121	18112	14648	2774	1204	
21	19	7677	1209	5019	2923	1890	1040	4796	7110	11689	9612	2177	996	
22	20	8013	1122	5111	2451	2054	2118	4061	5113	20919	7111	1810	1019	
23	21	12146	1119	3397	1016	4251	6138	3478	5791	82716	5113	1712	1013	
24	22	7371	2084	4989	2016	15013	4916	3650	4744	64691	4361	1736	1019	
25	23	18468	4919	4170	1764	4011	4717	18718	14270	14072	16401	19101	4241	
26	24	16840	4168	3511	1719	5281	4118	17217	4247	19101	1300	1689	11941	
27	25	11702	1188	3019	1889	1940	2411	8796	4189	12880	2718	6130	8711	
28	26	8471	10824	2261	2170	5543	3020	6118	4018	10070	2411	9477	3719	
29	27	6902	2988	1716	3020	3790	2020	5481	20392	9128	2942	7874	3398	
30	28	1141	1078	2400	4617	4915	1296	1216	9617	1066	1038	6940	2961	
31	29	4419	1000	4864	3877	2511	4977	6716	1065	2730	18412	12991		
32	30	3172	11761	4878	3428	2801	4941	509	5981	1007	10891	7191		
33	31	2145	6482	4480	4480	4480	5051	5050	2460			5101		
34	Jan	954	718	1610	1719	1117	1991	1812	4058	1891	2198	1572	998	718
35	Mar	4018	1122	8979	14182	3402	1011	7940	13419	11919	1718	6175	7171	8108
36	May	17068	10824	56282	11371	15013	18165	17647	30518	82736	25118	6940	47176	91171
37	Daily mean flow values 2001													
38	Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	
39	1	4027	1415	408	297	4268	4205	3408	10124	1028	7890	2090	2090	
40	2	1190	218	1197	107	2467	1814	4474	11917	4917	7471	2948	1660	
41	3	1778	112	168	171	1811	1688	6177	10444	1048	1842	1148	1500	
42	4	2512	210	617	211	1503	2688	26135	10027	8399	7114	1087	1219	
43	5	2185	206	476	218	1117	13405	11378	9436	4730	8122	1028	1011	
44	6	2188	119	469	281	924	20182	42611	7908	1796	12072	1124	1212	
45	7	1409	1019	340	201	861	70072	29649	1629	1928	8113	1642	2219	
46	8	1412	1015	351	411	750	18887	24619	6771	4172	1041	1041	7811	
47	9	1223	615	354	466	798	12171	13007	9601	4071	5791	1019	16627	

- No es solo el almacenamiento, es el procesamiento

Base de datos, BD

- **BD**: colección organizada de datos, que
 - modela aspectos relevantes de la realidad
 - » clientes, llamadas, tarifas
 - da soporte a procesos de información
 - » encontrar 10% clientes con menor relación de gasto fijo / móvil
- **SGBD**: sistema gestor de BD
 - sistema software que gestiona los datos
- Las BDs se diseñan, los SGBDs se usan

Gestión de datos

- Gestión: almacenamiento, extracción, modificación, borrado, búsqueda, seguridad, integridad, compartición, ...
- Modelo relacional: los datos se describen como tablas relacionadas
 - ¿existen otros modelos?
 - ventajas / inconvenientes
- Existe una excesiva orientación hacia el almacenamiento

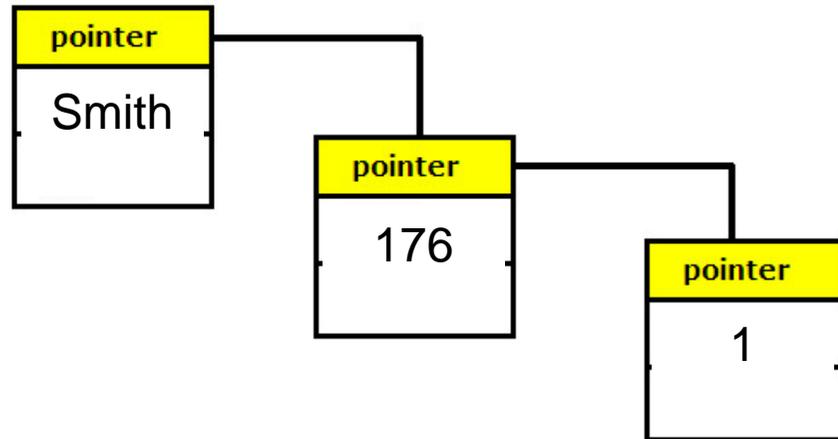
Almacenamiento vs BD



① hardware

- visión *demasiado* hardware
- en último término los datos tienen que ser almacenados así
- persistencia de los datos
– ¿qué se hace con los no-persistentes? –

...



② físico

- datos de bajo nivel pero con forma de **estructura de datos** (lista encadenada)
- archivo en memoria secundaria
- la estructura de datos me da una forma de acceso
- puedo servirme del sistema de archivos del sistema operativo

...

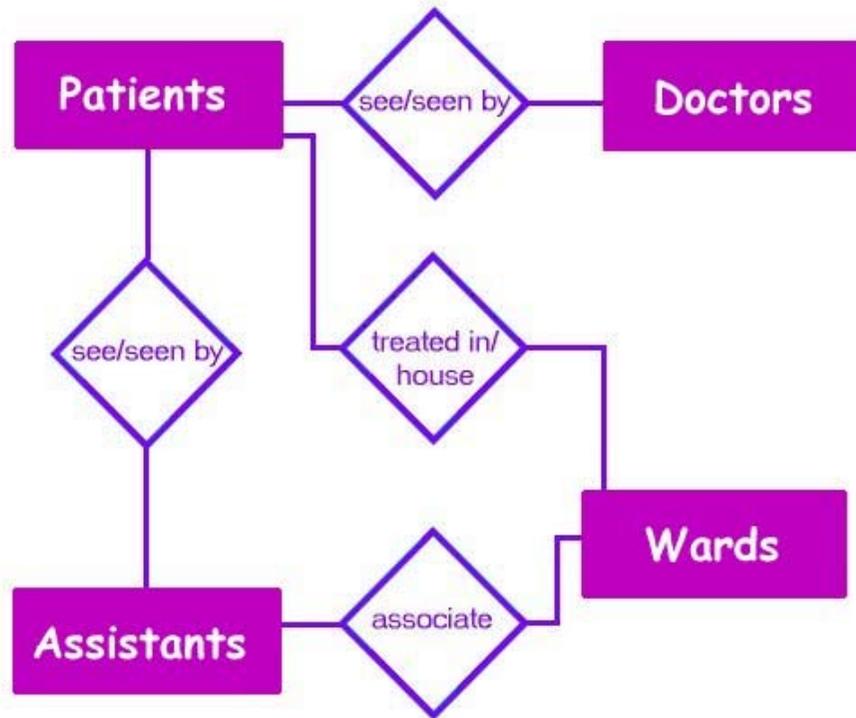
```
1 File: Hospital Patient Data
2 Format: Comma-Separated File
3
4 YPL-320, SMITH, m, 38, 176, 1, 124, 93
5 GLI-532, JOHNSON, m, 43, 163, 0, 109,
6 PNI-258, WILLIAMS, f, 38, 131, 0, 125
7 MIJ-579, JONES, f, 40, 133, 0, 117, 75
8 XLK-030, BROWN, f, 49, 119, 0, 122, 86
9 TFP-518, DAVIS, f, 46, 142, 0, 121, 76
10 LPD-746, MILLER, f, 33, 142, 1, 130, 8
11 ATA-945, WILSON, m, 40, 180, 0, 115, 8
12 VNL-702, MOORE, f, 38, 182, 0, 115, 76
```

añadir pruebas que se realizan a cada paciente

③ lógico

- comprensible por agrupación de columnas
- tiene estructura de tabla basada en **entidades, relaciones y propiedades**
- la tabla no es una estructura de datos, aunque el fichero secuencial sí que lo es
- las líneas 1-3 son *especiales*
- la 3 constituye el esquema de los datos
- problemas para almacenar todo en una tabla

...



④ conceptual

- modelado de los conceptos que describen una organización en bases a los datos que gestiona
- visión global de alto nivel, basada en **entidades, relaciones y propiedades**
- o en clasificadores, asociaciones y atributos
- no se ven las *instancias* de los datos
- confusión entre modelos/esquemas/diseños conceptuales y lógicos

Sistemas de archivos

- Estructura de almacenamiento
 - enormes volúmenes de datos
 - compartición o replicación
 - identificación / indexación de los datos
- Programas de acceso a los datos
 - complejidad de las consultas y actualizaciones
 - comprobaciones de integridad (en cada uno)
 - consistencia en accesos concurrentes
 - consistencia ante fallos
- Seguridad
 - directivas de seguridad para distintos accesos

Sistemas de BDs

- Las BDs dan respuesta completa y eficaz a todo lo anterior

1. Independencia lógica-física: ② | ③

- usuario trata los datos a nivel lógico-conceptual
- internamente se puede cambiar el nivel físico
 - el usuario vería los datos exactamente igual, manteniendo todas sus aplicaciones intactas
- la 'traducción' la hace automáticamente el SGBD, manteniendo los detalles ocultos
- facilidad de cambio y mantenimiento para el acceso eficiente a los datos

...

2. Integridad y seguridad de los datos

- el SGBD asegura las restricciones de integridad, en vez de los múltiples programas de acceso
 - » salario mínimo de un empleado
- controles de acceso a datos visibles para cada usuario

3. Centralización de los datos

- minimizar la redundancia; evitar inconsistencias

4. Acceso concurrente y recuperación

5. Reducción del tiempo de desarrollo y mantenimiento de aplicaciones

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

...

