

UF2214: Implementación y uso de una BD

Elaborado por: Ángela Mañas Martínez

Edición: 5.1

EDITORIAL ELEARNING S.L.

ISBN: 978-84-16360-90-1

No está permitida la reproducción total o parcial de esta obra bajo cualquiera de sus formas gráficas o audiovisuales sin la autorización previa y por escrito de los titulares del depósito legal.

Impreso en España - Printed in Spain

Presentación

Identificación de la unidad formativa

Bienvenido a la Unidad Formativa **UF2214: Implementación y uso de una BD**. Esta Unidad Formativa pertenece al **Módulo Formativo MF0966_3: Consulta y manipulación de información contenida en gestores de datos** que forma parte del Certificado de Profesionalidad **IFCD0211: Sistemas de gestión de información**, de la familia de **Informática y Comunicaciones**.

Presentación de los contenidos

La finalidad de esta unidad formativa es enseñar al alumno a extraer informaciones contenidas en gestores de datos de distinta tipología, utilizando herramientas y lenguajes de consulta y manipulación de la información, de acuerdo a especificaciones técnicas y funcionales.

Para ello, se analizará el diseño y los **métodos de acceso a una BD** y los lenguajes de consulta y extracción de datos. También se realizará análisis del lenguaje de consulta para BD relacionales.

UF2214: Implementación y uso de una BD

Objetivos del módulo formativo

Al finalizar este módulo formativo aprenderás a:

- Distinguir las características y funciones de los lenguajes y herramientas de consulta y extracción de información de los sistemas gestores de datos, teniendo en cuenta sus especificaciones técnicas y funcionales.
- Aplicar procedimientos de extracción y consulta de información en el sistema gestor de datos, utilizando lenguajes de consulta específicos según necesidades dadas.

Índice

UD1. Diseño de BD	9
1.1. Etapas del diseño de bases de datos.....	11
1.2. Teoría de la Normalización	207
1.2.1. Primera forma normal	212
1.2.2. Segunda forma normal.....	214
1.2.3. Tercera forma normal.....	215
1.2.4. Forma normal de Boyce-Codd	217
1.2.5. Cuarta forma normal.....	219
1.2.6. Quinta forma normal. Forma normal de Boyce-Codd .	221
1.3. Aplicación de la Teoría de la Normalización al diseño de BD relacionales	222
1.4. Desnormalización de BD	229
UD2. Métodos de acceso a una BD	239
2.1. Implementación de los accesos por posición.....	241
2.2. Implementación de los accesos por valor.....	271

UF2214: Implementación y uso de una BD

2.2.1. Índices.....	272
2.2.2. Árboles B+	274
2.2.3. Dispersión	276
2.2.4. Índices agrupados	279
2.3. Implementación de los accesos por diversos valores.....	281
2.3.1. Implementación de los accesos directos.....	282
2.3.2. Implementación de los accesos secuenciales y mixtos.....	284
UD3. Lenguajes de consulta y extracción de datos	295
3.1. Características generales	297
3.2. SQL	298
3.3. OQL.....	356
3.4. JPQL.....	384
3.5. XMLQL.....	400
3.6. Otros.....	402
3.6.1. HTSQL	402
3.6.2. LINQL.....	404
UD4. Análisis del lenguaje de consulta para BD relacionales .	421
4.1. Sentencias de definición de datos	423
4.2. Sentencias de manipulación de datos	426
4.3. Sentencias de concesión y revocación de privilegios.....	431
4.4. Procedimientos almacenados.....	432
4.5. Disparadores	436

Índice

Glosario 445

Soluciones 447

Área: informática y comunicaciones

UD1

Diseño de BD

- 1.1. Etapas del diseño de bases de datos
- 1.2. Teoría de la Normalización
 - 1.2.1. Primera forma normal
 - 1.2.2. Segunda forma normal
 - 1.2.3. Tercera forma normal
 - 1.2.4. Forma normal de Boyce-Codd
 - 1.2.5. Cuarta forma normal
 - 1.2.6. Quinta forma normal. Forma normal de Boyce-Codd
- 1.3. Aplicación de la Teoría de la Normalización al diseño de BD relacionales
- 1.4. Desnormalización de BD

1.1. Etapas del diseño de bases de datos

Desde el punto de vista de la Informática entendemos que una base de datos es información agrupada o estructurada. Dicha información se organiza en conjuntos compartiendo al menos una característica. Los conjuntos se almacenan en discos que permiten el acceso directo. Junto a ellos se guardan programas que permiten la manipulación de los datos y SGBD. Las BD son herramientas fundamentales en la actual Sociedad del Conocimiento.



El Sistema de Gestión de Bases de Datos (SGBD) es una agrupación de programas informáticos muy especializados que sirven de canal de comunicación o de interfaz entre la base de datos, el usuario y las aplicaciones que la utilizan. Con los SGBD se definen, construyen y manipulan las BD (bases de datos). Gracias a ellos se almacenan las mismas y se puede acceder a los datos de forma rápida y organizada (ya se hayan organizado los datos de forma estructurada o desestructurada).



Antes de que aparecieran los conceptos actuales de BD la información se almacenaba en los inicios de la Informática en los llamados 'archivos planos'. Recibían este nombre porque carecían de estructura. Únicamente existían los campos, registros o filas y columnas. En los años 50 del siglo XX se utilizaban las cintas magnéticas. Consistía esta técnica en grabar la información en pistas sobre una banda plástica con un material magnetizado. Dicho material solía ser óxido de hierro o algún cromato. Fue en los años 60 del pasado siglo, con la llegada del concepto de DATA BASE, cuando se comienza a solucionar desde el punto de vista informático el acceso a la información. En esta década la información podía consultarse ya directamente, sin tener que conocer la ubicación exacta de los datos en el disco.

Las bases de datos de hoy son mucho más complejas que las primeras que aparecieron desde mediados del siglo XX pero también la tecnología avanza tan rápido que estar al día de todos los cambios en el universo de las bases de datos es una tarea difícil.

Consideramos que una base de datos es un conjunto de informaciones, que pertenecen al mismo contexto, almacenados de forma organizada para su posterior recuperación rápida y eficaz. Entonces, una biblioteca podría considerarse como aquella BD formada por textos impresos en papel, indexados y organizados para su consulta o rápida recuperación. Desde mediados del siglo XX por el rápido desarrollo tecnológico de la Informática y la Electrónica la mayoría de las bases de datos tienen formato digi-

tal. La característica principal es la inmediatez y es la solución para el almacenamiento de grandes cantidades de datos.

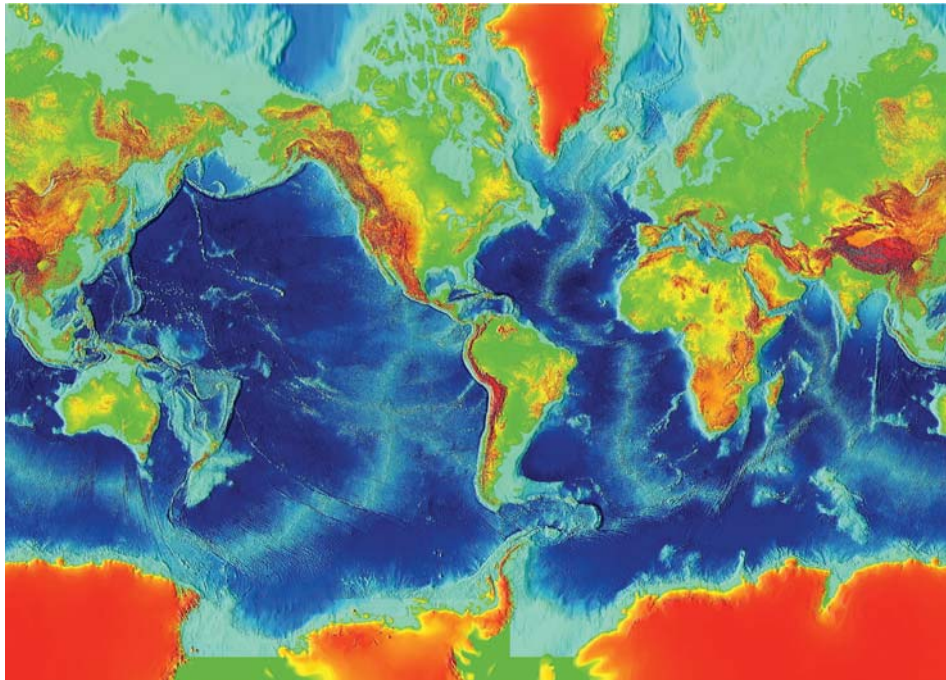


Imagen del relieve del planeta.

En el mundo globalizado en el que vivimos los datos carecen de fronteras. La cada vez mayor complejidad de las BD se acompaña de una tecnología que avanza por minutos. La puesta al día parece complicada. Sin embargo los mayores avances parten siempre de unas sólidas premisas básicas. Reconocerlas y haberlas estudiado abren la puerta del universo de las últimas tendencias tecnológicas.



Libros antiguos organizados.

En las primeras bibliotecas ya se organizaba la información de forma óptima y accesible. La división de los documentos por materias existió desde los inicios de la civilización. Primero en soportes de tablillas de barro, más tarde en papiros o códices de piel para terminar en la utilización del papel. La gran revolución de nuestros días viene de la mano del soporte digital que, aun ofreciendo posibilidades antes inimaginables, sigue fiel a unas premisas básicas que no han variado.

El mercado de bases de datos y de productos para la gestión, en las primera décadas del siglo XXI, excede con mucho los 12.000 millones de dólares al año (según datos de James R. Groff) y se ha duplicado en menos de una década. Ya se habla de mercado 'maduro' entre los fabricantes. El mayor beneficio en este sector viene de la mano de nuevas aplicaciones y nuevos productos como la arquitectura cliente/servidor o arquitecturas de sitios web, almacenes de datos y análisis de negocios basados en programas normalizados de gestión de bases de datos. Según este mismo autor la tecnología y los desarrollos relacionados con bases de datos seguirán proporcionando beneficios en las décadas venideras.

Está mundialmente aceptada la capacidad y utilidad de las bases de datos relacionales, cuyo formato se ha aceptado como la principal tecnología de procesamiento de datos empresarial. Han sido implementadas en casi todas las grandes corporaciones.

Como se ha dicho antes, las bases de datos de hoy son mucho más complejas que las primeras que aparecieron desde mediados del siglo XX pero también la tecnología avanza a gran velocidad.



El término 'DATA BASE' se escuchó por primera vez en 1963 en el Simposio de Santa Mónica sobre 'Optimización' organizado por la Universidad de California (EEUU). En él participaban alrededor de 250 matemáticos e ingenieros en un ambiente de colaboración científica auspiciado por la Rand Corporation.

Día a día, las bases de datos están teniendo una presencia cada vez más importante en nuestras vidas, de la que quizá no seamos del todo conscientes. Desde la consulta de nuestro saldo en el cajero automático; la disposición de nuestro dinero; las domiciliaciones de recibos o la compra en el supermercado se realizan con el uso de bases de datos. También por el cobro de un café por parte del camarero en nuestro local habitual, la lectura de la noticia más leída en periódico digital o la compra de un neumático en el taller. Todas estas actividades están relacionadas con bases de datos relacionales. Cuando las grandes, medianas y pequeñas empresas han establecido las bases de datos relacionales como marca y como estándar en el sector, hay una gran resistencia a cambiar esta norma.

Las **bases de datos relacionales** se han impuesto por su bajo coste, sencillez en su diseño e implementación y eficiencia en la recuperación de datos. Además la existencia de grandes fuerzas de ventas, relaciones establecidas con los clientes, además de acuerdos entre ellos han convertido a las bases de datos relacionales en las ganadoras.

Pero las bases de datos relacionales no son las únicas en el mercado y más allá del importante universo empresarial, puede tener una eficacia mayor otro tipo de estructura que no sea relacional. Para poder entender qué es una base de datos hay que entender su origen. La humanidad siempre ha querido conservar de forma organizada el conocimiento. Desde los orígenes de la civilización se ha sabido que la "información es poder".

La cultura más antigua, los sumerios (Mesopotamia, 3500 años antes de Cristo), trataba como clase social privilegiada a los escribas. Ellos eran los dueños de los datos. Eran los encargados de escribir sobre tablillas de arcilla y con escritura cuneiforme, las

transacciones mercantiles, cosechas, ventas de ganado. Para mantener el orden en un mismo 'registro', la tablilla dos comenzaba por la última frase de la tablilla uno. La tablilla tres, comenzaba por la última frase de la tablilla dos. Ya entonces, organizaban estos registros en cajas por temáticas, según cuenta Hipólito Escolar.

Más cerca, en el año 300 a.C, Alejandro Magno mandó construir en Egipto la Biblioteca de Alejandría, de uso público. Copió los papiros existentes en las bibliotecas privadas. Los manuscritos copiados llegaban a la ciudad egipcia de Alejandría por mar, desde todos los puntos de su conquista. De la misma manera que Google hoy intenta escanear todos los libros del mundo y ponerlos a disposición del mundo entero.

Y tanto ciudadanos romanos, como cristianos coptos y árabes disfrutaron de los títulos que comprendían toda la ciencia, la historia y la literatura de la Antigüedad durante siglos. La organización por temas fue cada vez más sofisticada.

Desde los orígenes de la Civilización, el hombre ha querido controlar y gestionar la información. La citada "La información es poder" puede parecer una frase del mundo cinematográfico del siglo XX. Sin embargo desde las primeras culturas que utilizaron el habla y la escritura se conocía esta verdad.

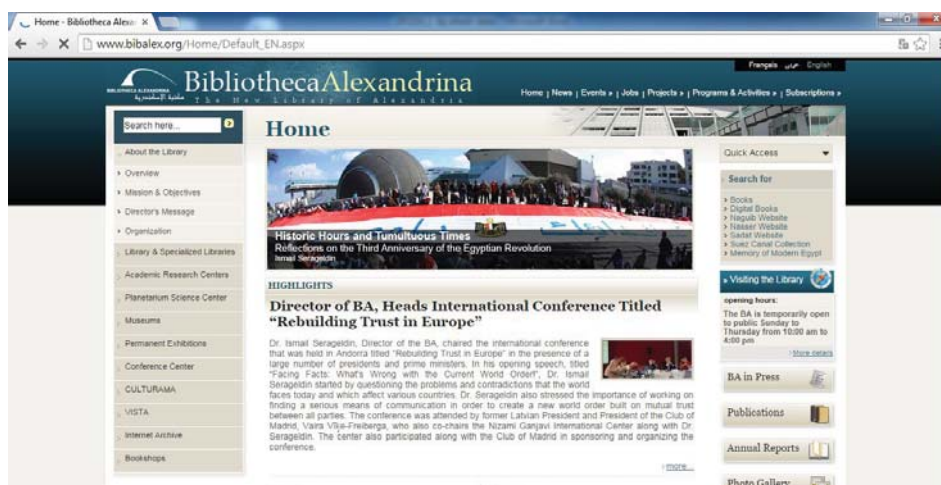


La clase social de los escribas sumerios, como ejemplo antropológicamente más antiguo, ya fue un estrato social privilegiado entonces.

Obviamente el acceso a los datos, a la información, fue durante muchos siglos privilegio de unos pocos.

La Biblioteca ideada por Alejandro Magno fue destruida sin que todavía se sepa con exactitud cómo ni con certeza por quién.

En 2002 la Unesco construyó la Nueva Biblioteca de Alejandría. Desde su interior podemos acceder a una ingente cantidad de información, a través de sus bases de datos bibliográficas (sobre libros), entre otras.



Web de Acceso a la Nueva Biblioteca de Alejandría, desde donde accedemos a una ingente cantidad de información.



Sabías que

La Biblioteca de Alejandría llegó a albergar hasta un millón de papiros con todo el conocimiento de la época. El papiro, que hasta entonces no había sido un material especialmente apreciado, se puso en valor. Hasta tal punto era la voracidad de la Biblioteca de Alejandro Magno que en varias ocasiones el papiro se agotó.

Bases de datos: origen y evolución histórica

Durante siglos, la organización de la información en las bibliotecas se fue haciendo más compleja. El estudio de cómo ordenar los documentos dentro de las cajas o el estudio de las cajas de libros existió durante el Imperio Griego y el Romano. En la Alta Edad Media y Baja Edad Media los sistemas de clasificación de materias y el Acceso a la cultura variaron. La Orden Benedictina fue la que más se dedicó a copiar a mano los libros sobre piel de oveja. Los primeros escribas sumerios pasaron a ser monjes benedictinos en la Edad Media. Trabajando en pupitres de madera copiaban letra a letra los manuscritos que, a veces, era decorado con oro. Gracias a ellos nos ha llegado el conocimiento matemático, literario y científico de la Antigüedad. Todo ese conocimiento simplemente ha cambiado de soporte y ahora lo tenemos en formato digital.

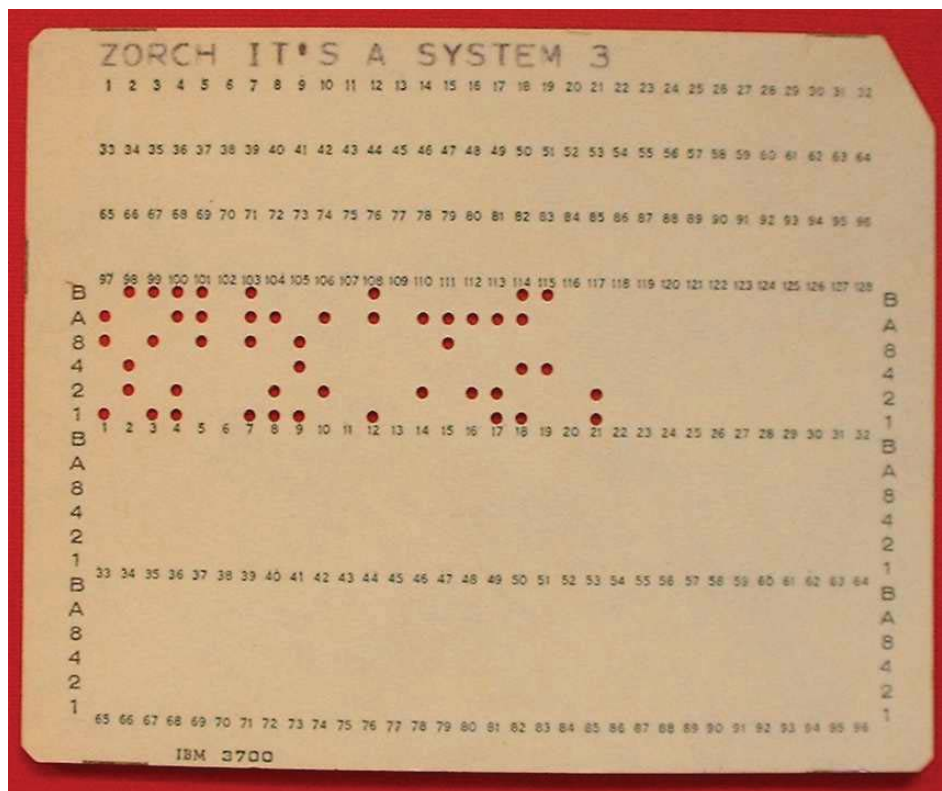
Durante los siglos XVI, XVII y XVIII el acceso a la lectura y a la educación era escaso y los libros siguieron considerándose como entretenimiento de aristócratas y reyes. Las bibliotecas fueron lugares poco accesibles para la mayoría. La forma de organizar el conocimiento siguió evolucionando. Hasta la invención de la imprenta (Johannes Gutenberg, 1398) los libros se escribían a mano. No era raro encontrarlos encadenados a la pared en las escasas bibliotecas de la realeza, aristocracia y el clero.



Los libros, o los datos que diríamos hoy, eran un objeto de lujo, inalcanzables para la mayoría.

Otro momento importante en la organización del conocimiento fue la invención de la Clasificación Decimal Universal en el siglo XIX. El filósofo pedagogo y psicólogo estadounidense John Dewey (Burlington, Vermont, 20 de octubre de 1859-Nueva York 1 de junio de 1952) ideó el citado Sistema de Clasificación Decimal (CDU). Dewey otorgaba a cada área del conocimiento un número subdivisible. Así el área de conocimiento de cada de libro podía ser representado por un número.

Más tarde, en 1884 Herman Hollerith creó la máquina automática de tarjetas perforadas, siendo nombrado así el primer ingeniero estadístico de la historia. En esta época, los censos se realizaban de forma manual.



Tarjetas perforadas de Hollerith. En la imagen, tarjeta procedente del Museo de IBM.

Ya en el siglo XX y antes de que la gestión de datos se realizara a través de los SGBD, todos los datos se almacenaban de forma permanente en los sistemas informáticos. Los registros de nóminas y contabilidad se guardaban en archivos individuales. Una aplicación, normalmente proporcionada por el fabricante de la computadora, como parte del sistema operativo, hacía un seguimiento del nombre y la ubicación de cada archivo. Este gestor de archivos carecía básicamente de modelo de datos. No consideraba en absoluto la organización interna de los archivos. Para el gestor de archivos eran iguales dos documentos que contuvieran diferentes informaciones.

El conocimiento sobre el contenido de los archivos –y los datos que contienen y el modo en que esos datos están organizados- se incorporaba en los programas de aplicación que utilizaban el archivo.

En esta aplicación para nóminas cada uno de los programas de COBOL que procesaban el archivo maestro del empleado contenía una *descripción del* archivo que indicaba la disposición de los datos en el archivo.

De esta forma, si se modificaba la estructura de los datos (por ejemplo, si había que almacenar un elemento de datos adicional para cada empleado), había que modificar cada programa que tuviera acceso al archivo. Puesto que el número de archivos y de programas crecía con el tiempo, cada vez se dedicaba más esfuerzo del departamento de procesamiento de datos a mantener las aplicaciones existentes, en lugar de desarrollar unas nuevas. De ahí que hacia finales de los años 50 del siglo XX ya viera la necesidad de desarrollar SGBD. La idea subyacente era sencilla, sacar de cada programa la definición del contenido y la estructura de los datos y almacenarla, junto con los datos, en lo que fueron las primeras bases de datos.

Siglo XX

Década de los años 50 del siglo XX

Hollerith diseñó la primera máquina de tabuladores o censadora, basada en su primera máquina de tarjetas perforadas y perfeccionada. En los años 50 aparecen las cintas magnéticas para automatizar tareas de información y manejo de datos. Las nuevas industrias contaron entonces por primera vez con sistemas para servir la información que necesitaban las nuevas industrias.

Aparecieron los primeros ordenadores pero su precio los hacía inaccesibles. Con ellos fueron mecanizándose los primeros procesos de automatización de la información. En aquel momento solo se podía hacer de forma secuencial.