

Diplomatura en Bases de Datos

Centro de e-Learning SCEU UTN - BA.

Medrano 951 2do piso (1179) // Tel. +54 11 4867 7589 / Fax +54 11 4032 0148

www.sceu.frba.utn.edu.ar/e-learning



UTN.BA
UNIVERSIDAD TECNOLÓGICA NACIONAL
FACULTAD REGIONAL BUENOS AIRES

**Centro de
e-Learning**

p. 2

Módulo 1: Introducción y Fundamentos de Bases de Datos

Unidad 1: Introducción

Centro de e-Learning SCEU UTN - BA.

Medrano 951 2do piso (1179) // Tel. +54 11 4867 7589 / Fax +54 11 4032 0148

www.sceu.frba.utn.edu.ar/e-learning



Presentación:

En esta Unidad recorreremos la historia de las bases de datos para entender qué son y qué necesidad de negocios satisfacen. Vemos, también, cómo influyó en esa historia el aumento de la capacidad de almacenamiento de datos y también la creciente abundancia de información. Finalmente, repasamos el impacto del incremento de la velocidad de acceso así como de la creciente demanda de inmediatez por parte de los usuarios finales.



Objetivos:

Que los participantes:

- Entiendan el encuadre de las bases de datos dentro de una organización
- Tomen conocimiento de la evolución de las capacidades de las bases de datos desde su creación a la fecha
- Vean las diferencias entre las distintas aproximaciones al problema del almacenamiento y la recuperación de los datos.



Bloques temáticos:

1. ¿Qué es una base de datos?
2. Historia de las bases de datos.
3. Evolución de la capacidad de almacenamiento.
4. Evolución de las velocidades de acceso.
5. Evolución de los caudales de información.



Consignas para el aprendizaje colaborativo

En esta Unidad los participantes se encontrarán con diferentes tipos de actividades que, en el marco de los fundamentos del MEC*, los referenciarán a tres comunidades de aprendizaje, que pondremos en funcionamiento en esta instancia de formación, a los efectos de aprovecharlas pedagógicamente:

- Los foros proactivos asociados a cada una de las unidades.
- La Web 2.0.
- Los contextos de desempeño de los participantes.

Es importante que todos los participantes realicen algunas de las actividades sugeridas y compartan en los foros los resultados obtenidos.

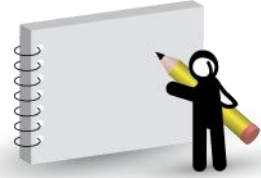
Además, también se propondrán reflexiones, notas especiales y vinculaciones a bibliografía y sitios web.

El carácter constructivista y colaborativo del MEC nos exige que todas las actividades realizadas por los participantes sean compartidas en los foros.

** El MEC es el modelo de E-learning colaborativo de nuestro Centro.*

Centro de e-Learning SCEU UTN - BA.

Medrano 951 2do piso (1179) // Tel. +54 11 4867 7589 / Fax +54 11 4032 0148
www.sceu.frba.utn.edu.ar/e-learning



Tomen nota:

Las actividades son opcionales y pueden realizarse en forma individual, pero siempre es deseable que se las realice en equipo, con la finalidad de estimular y favorecer el trabajo colaborativo y el aprendizaje entre pares. Tenga en cuenta que, si bien las actividades son opcionales, su realización es de vital importancia para el logro de los objetivos de aprendizaje de esta instancia de formación. Si su tiempo no le permite realizar todas las actividades, por lo menos realice alguna, es fundamental que lo haga. Si cada uno de los participantes realiza alguna, el foro, que es una instancia clave en este tipo de cursos, tendrá una actividad muy enriquecedora.

Asimismo, también tengan en cuenta cuando trabajen en la Web, que en ella hay de todo, cosas excelentes, muy buenas, buenas, regulares, malas y muy malas. Por eso, es necesario aplicar filtros críticos para que las investigaciones y búsquedas se encaminen a la excelencia. Si tienen dudas con alguno de los datos recolectados, no dejen de consultar al profesor-tutor. También aprovechen en el foro proactivo las opiniones de sus compañeros de curso y colegas.



1. ¿Qué es una base de datos?

Las organizaciones en general y las empresas en particular vienen generando crecientes volúmenes de datos.

Si pensamos en los registros que se tomaban hace 200 años nos encontramos con libros, cartas personales y poco más. Unas pocas organizaciones comenzaban a registrar sus movimientos de dinero y bienes.

El medio casi exclusivo que utilizaban era el papel y la forma de leer y buscar era por simple inspección visual.

La información estaba entonces contenida en libros los cuales se organizaban en bibliotecas.

Cuando, por el volumen de las organizaciones, la cantidad de información disponible hizo que las búsquedas y operaciones se hicieran impracticables se buscaron formas de organizar el contenido. Aparecieron las máquinas de contabilidad que constituyen uno de los precursores de nuestra industria informática.

El aumento de la cantidad de información generada incentivó la capacidad de los medios para guardar y procesar información.

La difusión de esos medios llevó los precios correspondientes a la baja lo que fomentó su uso y motivó el registro de más información que antes nunca se recogía.

Esta abundancia de datos no hizo sino subrayar la necesidad de optimizar la forma de operar con ellos. Las incipientes computadoras fueron la herramienta para lograr esta optimización.

En este curso nos proponemos recorrer las etapas de este proceso que todavía prestan servicios en activo combinando la descripción de las capacidades, ventajas y desventajas con las normas básicas de manejo de cada pieza de tecnología que vayamos considerando.



Al mismo tiempo, a medida que distintos conceptos vayan apareciendo los vamos a tratar por medio de "Notas conceptuales" que permitan incorporar la información dentro de su contexto temático e histórico.

Vamos a recorrer:

- Bases de datos relacionales
- Bases de datos de objetos
- Bases de datos documentales
- Bases de datos en memoria
- Bases de datos distribuidas
- Bases de datos orientadas a Big Data

En cada caso nos proponemos tanto la descripción como la aplicación de las herramientas básicas de manera que los participantes se puedan hacer una idea de la potencialidad y limitaciones de cada caso.

¿Cómo ocurrió todo?



2. Historia de las bases de datos

1884 - 1960

Herman Hollerith comenzó en 1884 a trabajar en el diseño de una maquina tabuladora o "censadora", basada en tarjetas perforadas.

Hacia 1950 se da origen a las cintas magnéticas, para automatizar la información y hacer respaldos, Y a través de este mecanismo se agilizó tanto el almacenamiento como la recuperación de los datos.

1960-1970

En la década de los 60, los precios de las computadoras empezaron a ser accesibles para las organizaciones de gran porte. También aparecen las primeras unidades de discos de almacenamiento que son muy superiores a las cintas cuando se trata de acceder a la información en forma aleatoria.

Nota conceptual:

Las cintas privilegian el devolver la información en el mismo orden en que fue grabada. A esto se llama acceso secuencial. Los discos, si bien son buenos para el acceso secuencial presentan mayores ventajas comparados con las cintas cuando se trata de acceder a distintas piezas de información sin seguir el orden en el que fue grabada.

En esta época también empezaron las primeras generaciones de bases de datos de red y las bases de datos jerárquicas.

Centro de e-Learning SCEU UTN - BA.

Medrano 951 2do piso (1179) // Tel. +54 11 4867 7589 / Fax +54 11 4032 0148

www.sceu.frba.utn.edu.ar/e-learning



Nota Conceptual:

Una base de datos jerárquica podemos verla como un caso particular de la base de datos en red. La particularidad del modelo jerárquico es que a cada nodo le pueden corresponder de 0 a N hijos pero a cada nodo hijo le corresponde un sólo padre

Durante este tiempo también se unieron IBM y American Airlines para crear SABRES, un aplicativo para controlar las reservas de vuelos, la información de los pasajeros y las transacciones.

Charles Bachman creó un nuevo tipo de bases de datos y esto permitió la creación de un estándar en los sistemas de bases de datos gracias a invención de nuevos lenguajes de sistemas de información como IDS.

Hacia 1960 no existía dentro de muchas empresas mecanismos que permitieran grabar, consultar y alterar la información de manera automática.

También dentro de esta década aparece por primera vez el término "Base de Datos" acuñado en un simposio celebrado en California en 1963.

Nota Conceptual:

Base de datos es un conjunto de información relacionada que se encuentra agrupada o estructurada.

En 1969 Edgar F. Codd acuña los conceptos de:

Derivability o capacidad para obtener un elemento de información.

Redundancy o redundancia.

Consistency o consistencia.

Centro de e-Learning SCEU UTN - BA.

Medrano 951 2do piso (1179) // Tel. +54 11 4867 7589 / Fax +54 11 4032 0148

www.sceu.frba.utn.edu.ar/e-learning



También se da cuenta que existen dos visiones posibles sobre la estructura interna de una base de datos:

- **Modelo físico:** que se relaciona con donde están físicamente grabados los datos en los medios de almacenaje
- **Modelo lógico:** que tiene que ver con cómo ve el usuario a los datos .

1970-1980

En los años 70 aparecen las primeras bases de datos relacionales como Ingres y System R.

Edgar Frank Codd, definió el modelo relacional a la par que publicó una serie de reglas para los sistemas de datos, Este hecho dio paso al nacimiento de la segunda generación de los Sistemas Gestores de Bases de Datos.

Lawrence J. Ellison, más conocido como Larry Ellison, basándose el trabajo de Edgar F. Codd, desarrolló el Relational Software System que dio lugar a Oracle.

Posteriormente se formalizó bajo el nombre de SQL (Structured Query Language) como un lenguaje de consultas o lenguaje declarativo de acceso a bases de datos relacionales que permite efectuar consultas con el fin de recuperar información de interés de una base de datos.

1980-1990

A principios de los años ochenta comenzó el despegue de la comercialización de los sistemas relacionales, y SQL comenzó a ser el estándar de la industria, ya que las bases de datos relacionales con su sistema de tablas (compuesta por filas y columnas)



1990-2000

En la década de 1990 la investigación en bases de datos se encaminó hacia las bases de datos orientadas a objetos.

También se desarrollaron herramientas como Excel y Access del paquete de Microsoft Office. Excel no alcanza a cumplir con las exigencias de una base de datos relacional aunque Access si lo hace. Algunos los consideran como el inicio de las bases de datos orientadas a objetos.

El desarrollo de la WWW presentará un doble papel respecto de las bases de datos. Por una parte las páginas web consumirán recursos de bases de datos para brindar información y por otro lado se convertirán en plataformas desde las cuales realizarles consultas.

2000-

Hoy, las tres grandes compañías que dominan el mercado de las bases de datos son IBM, Microsoft y Oracle.

Por su parte, en el campo de internet, la compañía que genera gran cantidad de información es Google. En su necesidad de tener almacenada la información que necesita su famoso buscador para funcionar se ha visto obligada a desbordar las capacidades (cada vez mayores) que hoy ofrecen las bases de datos relacionales e incursionar en lo que se vino a llamar Big Data mediante su producto Hadoop (entre otros)

Distintos tipos de necesidades han llevado al desarrollo y difusión de otros modelos como la base de datos orientada a documentos (MongoDB) o la base de datos orientada a grafos como Neo4j



3. Evolución de la capacidad de almacenamiento

Tenemos que distinguir dos tipos de almacenamiento. El almacenamiento temporario que se da típicamente en la RAM de una computadora y el permanente que se da en cintas, discos, CDs u otros medios.

La RAM es típicamente volátil y se borra cuando se apaga el dispositivo. Los discos tienden a tener un carácter permanente y la información no se borra con el apagado de los equipos.

A la hora de trazar el recorrido histórico de la capacidad de almacenamiento deberemos tener en cuenta ambas dimensiones:

al principio las discos eran grandes:



Centro de e-Learning SCEU UTN - BA.

Medrano 951 2do piso (1179) // Tel. +54 11 4867 7589 / Fax +54 11 4032 0148

www.sceu.frba.utn.edu.ar/e-learning



fuelle: wikipedia https://en.wikipedia.org/wiki/File:BRL61-IBM_305_RAMAC.jpeg

Este disco rígido era un componente característico de la IBM RAMAC 305 (Sistema de Contabilidad con Memoria de Acceso Aleatorio, que fue el primer ordenador de venta libre que usaba un disco duro de cabeza móvil (disco magnético).

Fue presentado en septiembre de 1956 y constaba de 40 a 50 platos con capacidad para tanto lectura como escritura mediante la cabeza de un solo brazo que se movía arriba y abajo de la pila de discos duros magnéticos.

Esos 50 discos podían almacenar unos 5 millones de caracteres, esto equivaldría hoy a 5 Mb aunque, nosotros usamos hoy bytes de 8 bits y en aquel momento se usaban de 7 bits.

La tasa de transferencia (throughput) era de 8.800 bytes por segundo.

Para mantener andando este monstruo se usaban dos refrigeradores. El costo era de U\$S 50.000 con lo que cada Mb necesitaba de unos U\$S 10.000.



Fuente: Imágenes de Google:

<https://www.google.com.ar/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwjCuZfBxe3bAhWGIJAKHbZ6APUQjRx6BAgBEAU&url=http%3A%2F%2Freyes.svetlian.com%2Fblog%2Fdiscos-duros.htm&psig=AOvVaw2qPQd3pTh3OHZBe8Z-6Jly&ust=1529972408732379>

En 1961, IBM crea cabezales de discos duros que permanecen flotando sobre un colchón de aire. Esto permite una mayor cercanía al disco lo que terminará facilitando una mayor compresión de los datos.

La Byrant Computer 4240 ya venía con un disco duro de 90MB. Con mayor integración se pudo bajar la cantidad de discos a 24 pero hubo de aumentar el diámetro de cada uno a poco menos de un metro (99.1 cm).

IBM presentó en 1962 los 1301 Advanced Disk File con tenían 25 discos de 60 cm de diámetro y capacidad de 28MB. Algo después, en 1963 llegó el 1311, con 6 platos de 28 cm de diámetro con una capacidad y 2.69 Mb.

Centro de e-Learning SCEU UTN - BA.

Medrano 951 2do piso (1179) // Tel. +54 11 4867 7589 / Fax +54 11 4032 0148
www.sceu.frba.utn.edu.ar/e-learning



Estos fueron los primeros modelos de discos podían ser removidos y cambiados. El 1311 fué tan exitoso que siguió vigente hasta mediados de los 70:



Fuente: Imágenes de Google:

https://www.google.com.ar/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwjsrpasxu3bAhXFhJAKHWI9Cj0QjRx6BAgBEAU&url=https%3A%2F%2Fes.wikipedia.org%2Fwiki%2FPrimeros_discos_IBM&psig=AOvVaw3DqCxpZK2vEQMTJubG9usc&ust=1529972608927878

La integración avanzaba y con eso las unidades se hacían cada vez más pequeñas y, a la vez, más poderosas. En 1965, se produjo el lanzamiento del 2310 Ramkit, con un disco único de 28 cm y una capacidad de 1MB por parte de quién se convertiría luego en Western Digital.



Fuente:

https://www.google.com.ar/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwj98NO9i-_bAhWMD5AKHYKgBwoQjRx6BAgBEAU&url=http%3A%2F%2Fwww.computerhistory.org%2Fcollections%2Fcatalog%2FL2006.2.6&psig=AOvVaw2IN7DtFylGSxg-0jzcH8mQ&ust=1530025550674154

IBM pugnaba por el liderazgo. En 1973 introdujo el modelo 3340, que es podríamos considerar como el padre del disco duro moderno. Esta unidad tenía dos ejes independientes, uno permanente y el otro desmontable. Cada uno tenía una capacidad de 30 MB.

Por este motivo muchos se han referido al 3340 como "30-30" que evocaba al rifle Winchester. Tal fué la difusión que alcanzó este modelo que el nombre del rifle se le pegó

Centro de e-Learning SCEU UTN - BA.

Medrano 951 2do piso (1179) // Tel. +54 11 4867 7589 / Fax +54 11 4032 0148

www.sceu.frba.utn.edu.ar/e-learning



al disco rígido y todavía, en algunos ambientes se habla del "winchester" para referirse a un disco rígido.

El 3340 usaba sellado interno y mejoró la tecnología denominada como "cojinete de aire", el disco Winchester con lo que reducía aún más la distancia de la cabeza a la superficie del disco.



Fuente: <https://commons.wikimedia.org/wiki/File:IBM3380DiskDriveModule.agr.jpg>

A principios de 1980 IBM lanzó 3380. Otro fuerte jugador de este mercado, Seagate, introdujo la primer unidad de 5.25": el ST507 con cuatro cabezas y 5 Mb.

IBM dejó pasar este primer modelo de Seagate pero tomó el siguiente, el ST-412, que tenía 10MB, para incorporarlo a la IBM PC/XT.



Fuente:

https://www.google.com.ar/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwiJtNW2je_bAhUEHpAKHWw6CaUQjRx6BAgBEAU&url=https%3A%2F%2Fwww.pcmag.com%2Farticle%2F0%2C2817%2C2134684%2C00.asp&psig=AOvVa w22M3om79rxa1H8yABsFKh0&ust=1530026025504571

En 1983 se introdujo el primer disco duro de 3.5", por parte de Rodime. Tenía 2 platos y almacenaba hasta 10MB. El cambio en tamaño era importante, especialmente para los que soñaban con que las computadoras pudieran llevarse.

Las mejoras se fueron acelerando. El IBM 3380 como el primer disco duro "hogareño" hogareño que podía almacenar hasta 1 GB de datos.

En 1981 Western Digital propuso el primer disco basado en un solo chip.

En 1988 Prairie Tek introdujo el primer disco de 2.5" que se volvería estándar en las computadoras portátiles.

Centro de e-Learning SCEU UTN - BA.

Medrano 951 2do piso (1179) // Tel. +54 11 4867 7589 / Fax +54 11 4032 0148

www.sceu.frba.utn.edu.ar/e-learning



4. Evolución de las velocidades de acceso

A medida que fueron evolucionando los modelos las velocidades a las cuales podían consultarse los datos fueron aumentando. Lo podemos resumir en esta tabla:

Modelo	Velocidad de acceso	
	Tiempo de búsqueda	Secuencial
IBM 350		8,8 Kb/s
IBM 1301	180ms	90 Kb/s
IBM 2310	85 ms	156 Kb/s
IBM 3340	25 ms	885 Kb/s
Seagate ST 412	85 ms	5000 Kb/s
IBM 3380	16 ms	3000 Kb/s
Hitachi 7K1000	20 ms	1284 Kb/s

Nota conceptual:

Tiempo de búsqueda:

Es el tiempo que tarda el disco rígido en devolver un dato en una posición al azar. Viene determinado por la velocidad de rotación del disco y la velocidad mecánica del brazo que mueve la cabeza lectora. Este valor refleja la agilidad del disco para realizar consultas pequeñas e independientes.

Tasa de transferencia:

Es la cantidad de datos por segundo que, en promedio, es capaz de leer un disco. Este valor refleja la capacidad de un disco para volcar información.



5. Evolución de los caudales de información

En 2011 se publicó un estudio en la revista Science donde se pretendía cuantificar la cantidad de información generada y almacenada en el mundo. Ese mismo año Eric Schmidt, CEO de Google, comentó que la Humanidad había creado hasta 2003 unos 5 Exabytes. Según esa fuente, el mismo volumen generado por toda la historia hasta 2003 se acumulaba en 2011 cada 2 días.

Según el mencionado artículo de Science el cambio es abrumador.

Para mayor claridad vamos a reflejarlo como una tabla:

Tiempo	Información
Hasta el 2003	5 Exabytes
Del 2003 al 2007	290 Exabytes
Durante el 2011	600 Exabytes

Entre algunas de ellas destacan la cantidad de información generada por la humanidad hasta el año 2007 que la estiman en 295 exabytes, aumentando en 2011 a 600 exabytes, o lo que es lo mismo un trillón de bytes, que es la capacidad que pueden contener un millón de ordenadores de sobremesa actuales. El estudio también nos dice que, la tecnología digital domina claramente sobre la analógicas puesto que desde el 2007, el 99,9% de la información generada era en formato digital, o al contrario, que sólo el 0,007% de la información del planeta está en papel.

El pasado año la empresa DOMO recogía en una infografía muchos datos interesantes respecto al uso que los internautas hacen de la Red en cada minuto que pasa. Viéndolo así, podemos apreciar las magnitudes que hay en cuanto a la creación de contenido y el tráfico de información en el mundo.

Cada minuto que pasa, los 2.700 millones de personas con acceso a Internet que se calcula que hay actualmente en el mundo envían más de 200 millones de correos electrónicos, realizan 2 millones de consultas a Google, suben 48 horas de vídeo a YouTube, escriben más de 100.000 mensajes en Twitter, publican casi 30.000 nuevos artículos en sitios como Tumblr o WordPress y suben más de 6.000 fotografías a Instagram y Flickr:

Centro de e-Learning SCEU UTN - BA.

Medrano 951 2do piso (1179) // Tel. +54 11 4867 7589 / Fax +54 11 4032 0148

www.sceu.frba.utn.edu.ar/e-learning

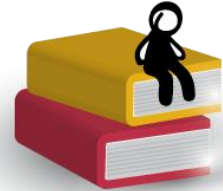


Fuente:

https://www.google.com.ar/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwi-1drTju_bAhXKFJAKHfKgCeQQjRx6BAGBEAU&url=https%3A%2F%2Fwww.domo.com%2Flearn%2Fdata-never-sleeps-2&psig=AOvVaw01JP4hBQ2wO6ltcr2KX3Va&ust=1530026394034776

Centro de e-Learning SCEU UTN - BA.

Medrano 951 2do piso (1179) // Tel. +54 11 4867 7589 / Fax +54 11 4032 0148
www.sceu.frba.utn.edu.ar/e-learning



Bibliografía utilizada y sugerida

Bourne, Charles P. A History of Online Information Services 1963-1976, Mit Press, USA, Massachusetts, Agosto 2003.

Date, C. J. An Introduction to Database Systems, Pearson Educación, Mexico, 2001

Hillbert Martin, The World's Technological Capacity to Store, Communicate, and Compute Information, Science Review, February 2011.

Mulheren, Ed, Database History, Outskirts Press, Incorporated, USA, Parker, Colorado, Junio 2006.



Lo que vimos:

En esta Unidad vimos qué es una base de datos, como surgió para cubrir necesidades de negocios concretas y cómo evolucionaron los discos rígidos que guardan los datos en cuanto a capacidad de almacenamiento, así como en cuanto a capacidad de transferencia de información.



Lo que viene:

En la próxima Unidad recorreremos el modelo relacional, entendiendo por qué surgen y cómo se usan las formas normales. Luego, pasamos revista a los distintos objetos que viven dentro de una base de datos. Finalmente, nos introducimos en la estructura interna de almacenamiento de información, para entender cómo impactan esas decisiones en la velocidad con que resuelven las peticiones que se les envían.

