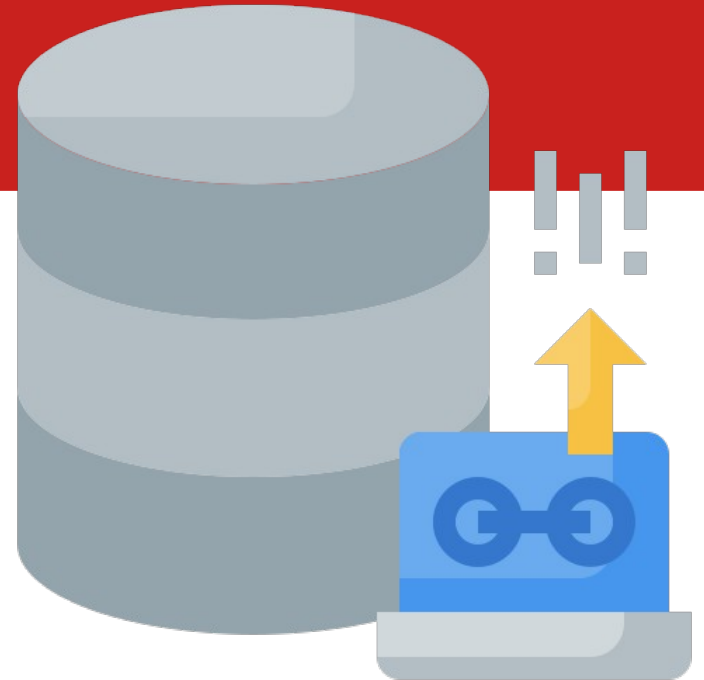


Bases de Datos I

Unidad I

Introducción a las Bases de Datos



Los orígenes de las bases de datos

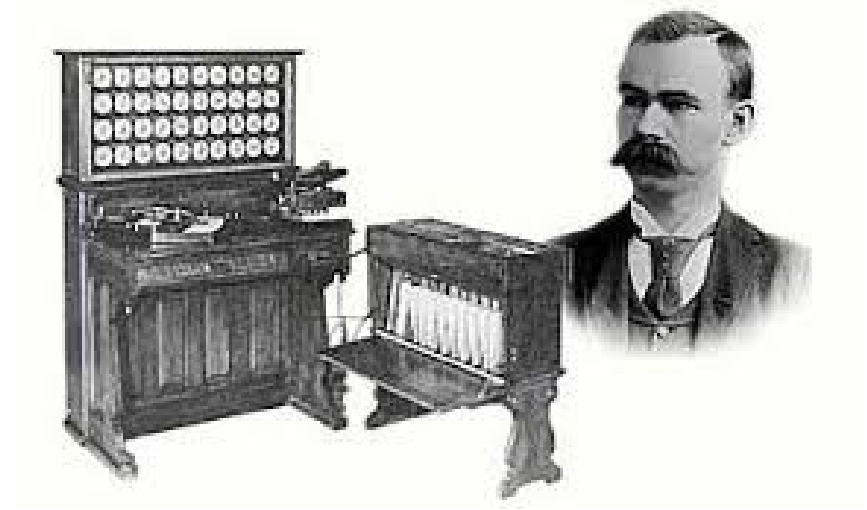
En la antigüedad ya existían bibliotecas y toda clase de registros.

- **Se utilizaban para recoger información sobre las cosechas y censos.**
- **No se contaba con la ayuda de máquinas que pudiesen reemplazar el trabajo manual.**

Los orígenes de las bases de datos

En 1884 Herman Hollerith creó la máquina automática de tarjetas perforadas.

- **En esta época, los censos se realizaban de forma manual.**
- **Hollerith desarrolló una máquina tabuladora o censadora, basada en tarjetas perforadas.**



Los orígenes de las bases de datos

- **En la década del '50 comenzaron a utilizarse las cintas magnéticas para el almacenamiento de datos.**
- **Solo permitían el acceso secuencial.**
- **El acceso y la búsqueda eran extremadamente lentos.**
- **Procesamiento por lotes.**

Los orígenes de las bases de datos

- **En la década del '60 se reduce el precio de las computadoras, por lo cual podían ser adquiridas por grandes empresas.**
- **El uso de cintas magnéticas es reemplazado por el uso de los primeros discos magnéticos.**
- **Permitían el acceso directo o aleatorio.**

Los orígenes de las bases de datos

- **Hasta este momento se utilizaban archivos para almacenar datos.**
- **Se escribían programas para almacenar y recuperar datos de los archivos.**
- **Esto daba lugar a numerosos problemas.**

Uso de archivos - Problemas (I)

Dependencia de los datos física-lógica

- **La estructura física de los datos (definición de archivos y registros) se encuentra codificada en los programas de aplicación.**
- **Cualquier cambio en la estructura implica identificar, modificar y probar todos los programas que manipulan esos archivos.**

Ejemplo

```
record tpersona  
  apellido:string[50];  
  nombre:string[50];  
  direccion:string[100];  
  telefono:integer;  
end;  
  
...  
personas:file of tpersona;  
  
...
```


Uso de archivos - Problemas (II)

Redundancia e inconsistencia

- **Los archivos son creados por distintos programas y van cambiando a lo largo del tiempo.**
- **El mismo dato puede tener distintos formatos y estar duplicado en varios sitios.**
- **La redundancia trae consigo la inconsistencia.**

Ejemplo

```
record tcliente  
    apellido:string[50];  
    nombre:string[50];  
end;
```

...

```
record templeado  
    apellido:string[60];  
    nombre:string[70];  
end;
```

...

Uso de archivos - Problemas (III)

Separación y aislamiento de los datos

- **Al estar repartidos en varios archivos, y tener diferentes formatos, es difícil escribir nuevos programas que aseguren la manipulación de los datos correctos.**
- **Antes se deberían sincronizar todos los archivos para que los datos coincidiesen.**

Ejemplo

```
record tpersona
```

```
...
```

```
direccion:string[100];
```

```
telefono:integer;
```

```
end;
```

```
...
```

```
record templeado
```

```
...
```

```
direccion:string[150];
```

```
telefono:real;
```

```
end;
```

Uso de archivos - Problemas (IV)

Dificultad para tener acceso a los datos

- **Cada vez que se necesite una consulta que no fue prevista en el inicio implica la necesidad de codificar el programa de aplicación necesario.**
- **Los entornos de procesamiento de archivos, no permiten recuperar los datos de una forma conveniente y eficiente.**

Ejemplo

- **Inicialmente**

- Necesitamos todas las personas, ordenadas por apellido y nombre.
- Necesitamos todas las personas, ordenadas por dirección.

- **Más tarde**

- Y ahora necesitamos los mayores de edad, que tengan teléfono y estén ordenados por nombre y apellido.
- Y necesitamos...



Uso de archivos - Problemas (V)

Dependencia de la estructura del archivo con el lenguaje de programación

- **La incompatibilidad entre archivos generados por distintos lenguajes hace que los datos sean difíciles de procesar.**

Ejemplo

- ¿Son lo mismo en PASCAL y C?

```
packed record tpersona
```

```
  apellido:string[50];
```

```
  nombre:string[50];
```

```
  telefono:integer;
```

```
end;
```

```
...
```

```
struct tpersona{
```

```
  char apellido[50];
```

```
  char nombre[50];
```

```
  int telefono;
```

```
};
```


Uso de archivos - Problemas (VI)

Problemas en la seguridad de los datos

- **Resulta difícil implantar restricciones de seguridad pues las aplicaciones se van añadiendo al sistema según se van necesitando.**
- **Los sistemas de gestión de archivos no permiten implementar mecanismos de seguridad de granularidad fina.**

Uso de archivos - Problemas (VII)

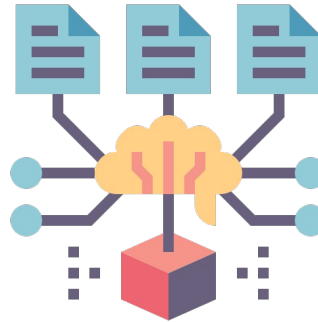
Problemas de integridad de datos

- **Los valores almacenados en los archivos deben cumplir con restricciones de consistencia.**
- **El problema se complica cuando existen restricciones que implican varios datos en distintos archivos.**

Orígenes de las bases de datos

- **Hace falta algo que solucione estos problemas y otros.**
- **La necesidad de almacenar y recuperar datos era cada vez más apremiante y necesaria.**
- **¿Qué es lo que hace falta?**

¡SOFTWARE!



Orígenes de las bases de datos

- **En 1961 Charles Bachman diseña el primer SGDB: IDS (Integrated Data Store) en General Electric.**
- **En 1973 Bachman se convirtió en la octava persona en ganar el premio Turing ACM.**



¿Qué es un SGBD?

- **SGBD (Sistema de Gestión de Base de Datos) o en inglés DBMS (Database Management System).**
- **Conjunto de programas que permiten a los usuarios crear, manipular, administrar y mantener una base de datos.**

ORACLE®

Informix®

Microsoft®
SQL Server®

MySQL®



PostgreSQL

¿Qué nos provee un SGBD?

- **Permite a los usuarios definir, crear la base de datos mediante el lenguaje de definición de datos (DDL).**
- **Permite la inserción, actualización, eliminación y consulta de datos mediante el lenguaje de manejo de datos (DML).**
- **Permite evitar la redundancia (datos repetidos).**

¿Qué nos provee un SGBD?

- **Permite evitar inconsistencias (que distintas copias de un dato tengan valores distintos).**
- **Permite evitar que usuarios no autorizados accedan a los datos (seguridad).**
- **Permite proteger los datos contra valores no permitidos (integridad).**
- **Permite concurrencia.**

¿Qué es una base de datos?

- **El término de bases de datos fue escuchado por primera vez en 1963, en un simposio celebrado en California, USA.**
- **Una base de datos se puede definir como un conjunto de información *relacionada* que se encuentra agrupada, *estructurada u organizada* según el o los modelos de datos que provee.**
- **Una Base de Datos es una colección de datos interrelacionados con un propósito específico vinculado a la resolución de un problema del mundo real. (Bertone y Thomas, 2011).**

A partir de la definición

Dominio de un problema

- **Deberemos seleccionar la información a representar de un dominio de un problema específico. Realizar el proceso de abstracción.**

Esquema de representación

- **Diseñar el esquema de representación de los datos (interrelación) que cumpla con ciertas características.**

¿Qué es un modelo de datos?

- **Un modelo de base de datos muestra la estructura lógica (no física) de la base de datos, incluidas las relaciones y limitaciones que determinan cómo se almacenan los datos y cómo se accede a ellos.**
- **Dentro del mundo de las bases de datos, existen distintos tipos de modelos de bases de datos.**
- **Un modelo de datos es un conjunto de herramientas conceptuales que permiten describir la información que es necesaria administrar para un Sistema de Información, las relaciones existentes entre estos datos, la semántica asociada y las restricciones de consistencia. (Bertone y Thomas, 2011)**

Modelo de datos

- **Los modelos de datos sirven para comprender más fácilmente los datos de una organización.**
- **Son medios para describir la realidad.**
- **Las personas que diseñan usan los modelos para construir esquemas, que sintetizan la realidad del problema.**

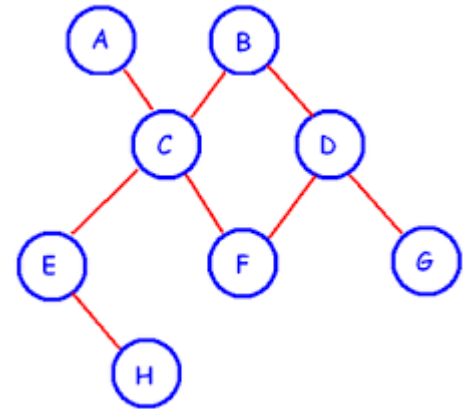
Relacionando conceptos

- **El modelado de datos es una parte esencial para el Diseño de Bases de Datos (BD). Las BD son una de las principales componentes de un Sistema de Información (SI).**
- **Un Sistema de Información es “... cualquier sistema usado para proveer información (incluyendo su procesamiento), para cualquier uso que pueda hacerse de ella...” (Langerfors, 1982).**
- **El sistema de datos es la parte estática que alimenta el sistema de información y se corresponde con los datos que se encuentran en la Base de Datos.**

Orígenes de las bases de datos

- **Con IDS nacía también uno de los primeros modelos lógicos de organizar los datos (modelo de datos):**

el Modelo de Red (simple).



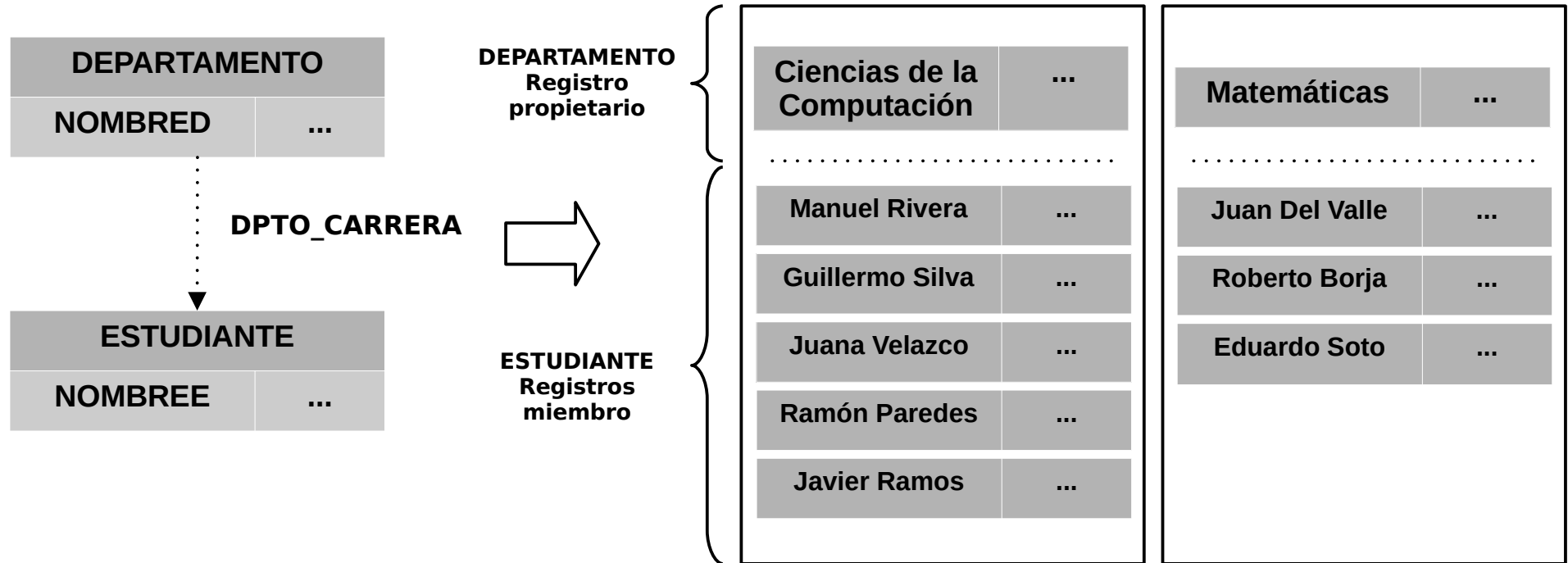
El modelo de red

- **Los datos se almacenan en registros.**
- **Los registros se clasifican en tipos de registros nombrados.**
- **Los registros están compuestos por elementos de información.**
- **Cada elemento de información tiene un nombre y un tipo de dato.**

El modelo de red

- **Para establecer relaciones entre registros se establecen tipos conjunto.**
- **Describen relaciones de uno a muchos.**
- **Tiene un nombre, un tipo de registro propietario (del lado uno) y un tipo de registro miembro (del lado muchos).**

El modelo de red



El modelo de red (DDL)

- **RECORD NAME IS DEPARTAMENTO**
NOMBRED TYPE IS CHARACTER 15
- **RECORD NAME IS ESTUDIANTE**
NOMBREE TYPE IS CHARACTER 20
- **SET NAME IS DEPTO_CARRERA**
OWNER IS DEPARTAMENTO
MEMBER IS ESTUDIANTE

El modelo de red (DML)

```
DEPARTAMENTO.NOMBRED := 'Investigacion';  
FIND ANY DEPARTAMENTO USING NOMBRED;  
IF(DB_STATUS = 0)THEN BEGIN  
    FIND FIRST ESTUDIANTE WITHIN DEPTO_CARRERA;  
    WHILE(DB_STATUS = 0)DO BEGIN  
        GET ESTUDIANTE;  
        WRITELN(ESTUDIANTE.NOMBREE);  
        FIND NEXT EMPLEADO WITHIN DEPTO_CARRERA;  
    END;  
END;
```

Orígenes de las bases de datos

- **En 1968 IBM introduce su producto IMS (Integrated Management System).**
- **Con IMS nacía otro modelo lógico de organizar los datos (modelo de datos):
el Modelo Jerárquico.**

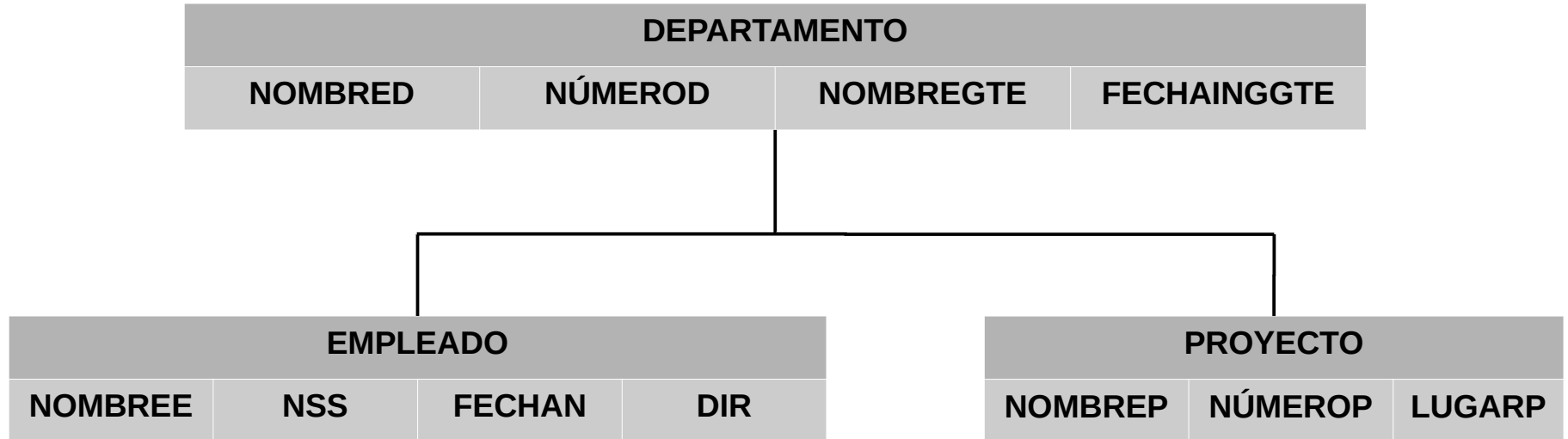
El modelo jerárquico

- **Los datos se almacenan en registros.**
- **Los registros se clasifican en tipos de registros nombrados.**
- **Los registros están compuestos por elementos de información.**
- **Cada elemento de información tiene un nombre y un tipo de dato.**

El modelo jerárquico

- **Para establecer relaciones entre registros se establecen vínculos uno a muchos de tipo padre-hijo.**
- **El registro del lado uno del vínculo se denomina registro padre.**
- **El registro del lado muchos del vínculo se denomina registro hijo.**

El modelo jerárquico



El modelo jerárquico (DDL)

- **SCHEMA NAME = EMPRESA**
HIERARCHIES = JERARQUIA
- **RECORD NAME = DEPARTAMENTO**
TYPE = ROOT OF JERARQUIA
NOMBRE CHARACTER 15
- **RECORD NAME = EMPLEADO**
PARENT = DEPARTAMENTO
NOMBREE CHARACTER 15

El modelo jerárquico (DML)

GET FIRST PATH

DEPARTAMENTO, EMPLEADO

WHERE NOMBRE = 'Investigacion';

WHILE(DB_STATUS = 0)DO BEGIN

WRITELN(EMPLEADO.NOMBREE);

GET NEXT EMPLEADO WITHIN PARENT;

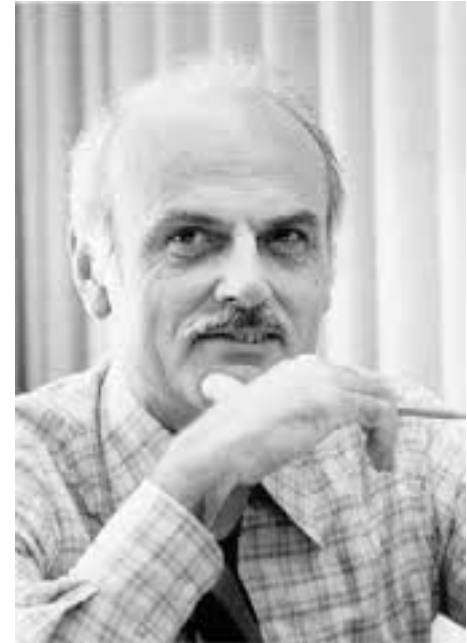
END;

Orígenes de las bases de datos

- **¿Qué tenían de malo estos modelos?**
- **Modelo Jerárquico**
 - **La realidad no siempre se estructura en forma jerárquica, esto dificulta el modelado.**
- **Ambos**
 - **Eran bases de datos navegacionales, las rutas a los datos quedan impregnadas en los programas lo que dificulta la modificación.**

Orígenes de las bases de datos

- **En 1970 Edgar Frank Codd, trabajando para el IBM Research Laboratory, publica:**
“A Relational Model of Data for Large Shared Data Banks”
- **Por su trabajo, ganó en 1981 el premio Turing ACM.**



El modelo relacional

- **A diferencia del modelo de redes y el modelo jerárquico, no se basaba en conceptos informáticos sino en conceptos matemáticos.**
- **El modelo se basa en las relaciones matemáticas y por consecuencia, en la teoría de conjuntos.**
- **Este es el modelo que vamos a trabajar a lo largo de toda la materia.**

Orígenes de las bases de datos

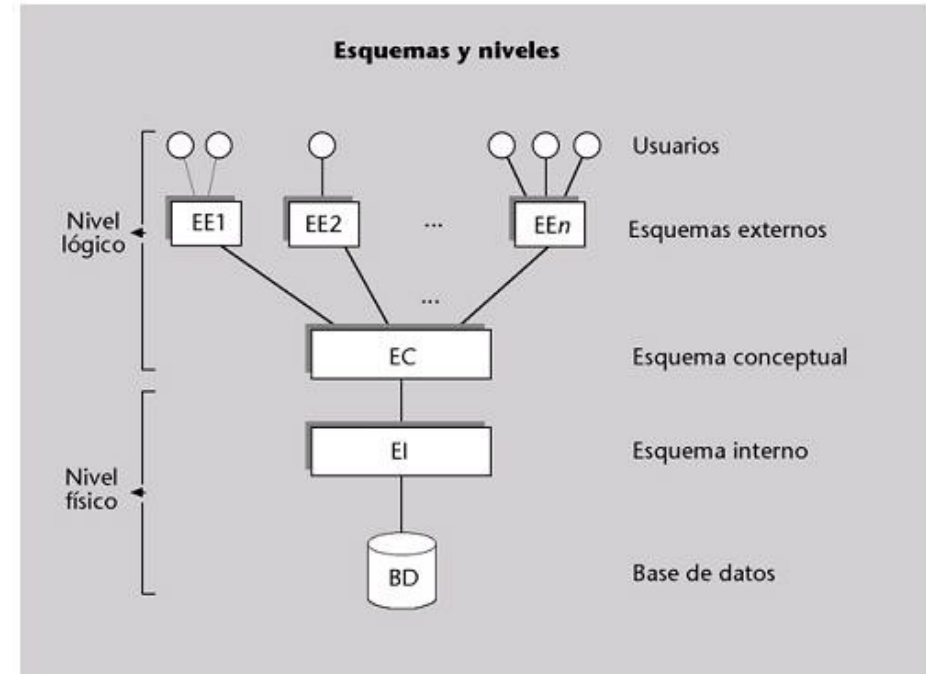
- **En 1970, a partir del trabajo de Codd, IBM comenzó a trabajar en prototipo de SGBD que soportara el modelo relacional.**
- **Esta implementación se llamó System R, la que luego se transformaría en SQL/DS y llegaría hasta nuestros días como DB2.**

Orígenes de las bases de datos

- **En 1973, miembros de la universidad de Berkeley comenzaron a trabajar en un SGBD llamado Ingres.**
- **Ese trabajo no solo dio lugar a ese SGBD, sino también a otros como Sybase, Informix y SQL Server.**
- **Los mismo miembros, luego crearon otro SGBD llamado Postgres, que luego se convertiría en PostgreSQL.**

Arquitectura de un SGBD

- En 1975, el comité ANSI-SPARC (American National Standard Institute - Standards Planning and Requirements Committee) propuso una arquitectura de tres niveles para los SGBD.



Fuente: https://virtual.itca.edu.sv/Mediadores/dbd/u2/11_arquitectura_del_sgbd.html

Modelo ANSI - SPARC

- **El objetivo principal era el de separar los usuarios y los programas de aplicación, de la base de datos física.**
- **Independencia lógica.**
- **Independencia física.**

Modelo ANSI - SPARC

Esquema externo o de visión:

- **Es el más cercano a los usuarios, es decir, es donde se describen varios esquemas externos o vistas de usuarios.**
- **Cada esquema describe la parte de la base de datos que interesa a un grupo de usuarios.**

Modelo ANSI - SPARC

Esquema conceptual:

- **Este esquema describe las entidades, atributos, relaciones, operaciones de los usuarios y restricciones, ocultando los detalles de las estructuras físicas de almacenamiento.**
- **Representa la información contenida en la base de datos.**

Modelo ANSI - SPARC

Esquema interno o físico:

- **Este esquema se especifica con un modelo físico y describe los detalles de cómo se almacenan físicamente los datos: los archivos que contienen la información, su organización, los métodos de acceso a los registros, los tipos de registros, la longitud, los campos que los componen, etcétera.**

Orígenes de las bases de datos

- **Para la década de 1980 existían numerosos SGBD relacionales, tanto para servidores o mainframes, como para computadoras personales.**
- **Ya en la década del 1990, este modelo terminó de establecerse como el modelo dominante en el mundo de las bases de datos.**

Orígenes de las bases de datos

- **Durante la década de 1990, nacieron también las bases de datos orientadas a objetos.**
- **En un primer momento se pensó que revolucionaría el mundo de las bases de datos al igual que lo hizo el modelo relacional.**
- **Esto no fue así, e incluso son raramente utilizadas.**

Orígenes de las bases de datos

- **El fracaso de las bases de datos orientadas a objetos se debe fundamentalmente a dos aspectos:**
 - **El mercado ya estaba totalmente acaparado por las bases de datos relacionales.**
 - **Implicaban un retroceso respecto al modelo relacional, dado que imponían nuevamente un modelo navegacional.**

En los últimos años

- **Con la explosión de internet, nacieron nuevas necesidades.**
- **Principalmente necesidades de flexibilidad, distribución y alta disponibilidad.**
- **Esto fue dando lugar a numerosos nuevos gestores de bases de datos y por consiguiente, nuevos modelos de datos.**
- **El modelo relacional sigue siendo el más utilizado para el desarrollo de software empresarial.**

En los últimos años

- **A los modelos de datos existentes se sumaron:**
 - **Orientado a documentos (Document Store)**
 - **Orientado a contenido (Content Store)**
 - **Orientado a columnas (Wide Column)**
 - **Orientado a series de tiempo (Time Series)**
 - **Orientado a clave-valor (Key-Value)**
 - **Orientado a grafos (Graph)**
 - **Orientado a buscadores (Search Engine)**
- **Se los agrupa bajo el término NoSQL (Not Only SQL).**

Hoy... NewSQL

- **Es difícil definir el término pero como indica su nombre, NewSQL es un nuevo paradigma de bases de datos relacionales que pretende añadir las capacidades NoSQL a las bases de datos relacionales.**
- **Capacidades como escalabilidad horizontal, sharding, datos semi-estructurados, etc.**

Bibliografía

- **Elmasri R. y Navathe Sh. (2000). Sistemas de Bases de Datos. Conceptos Fundamentales. Segunda Edición. Addison Wesley Longman de México, S. A.**
- **Introducción a las Bases de Datos. Fundamentos y Diseños (Bertone, Thomas)**