

# Fundamentos de Bases de Datos

## Tarea 1

### Conceptos básicos

Teresa Becerril Torres # de cuenta: 315045132  
Miguel Ángel Torres Sánchez # de cuenta: 315300442  
Nicole Romina Traschikoff García # de cuenta: 315164482  
Tania Michelle Rubí Rojas # de cuenta: 315121719

23 de agosto del 2019

## 1. Conceptos generales.

- a. ¿Por qué elegirías almacenar datos en un **sistema de base de datos** en lugar de simplemente almacenarlos utilizando el **sistema de archivos** de un sistema operativo? ¿En qué casos no tendría sentido utilizar un sistema de base de datos?

Porque en una base de datos bien diseñada se asegura la integridad de los datos y facilita trabajar con ellos a usuarios y desarrolladores. No tendría sentido usar una base de datos cuando los usuarios son pocos y cuando las cantidades de información no son grandes y no crecerán mucho o no existirá por largos periodos de tiempo.

- b. ¿Qué **ventajas** y **desventajas** encuentras al trabajar con una **base de datos**?

Las ventajas que proporciona trabajar con una base de datos son aquellas garantías que nos da su diseño y el sistema manejador de bases de datos, es decir, brinda chequeo de redundancia, operaciones básicas, integridad de datos y la posibilidad de seguir expandiendo un sistema robusto, además de poder recuperar la información de manera práctica y eficiente.

Por otro lado las desventajas principales pueden estar relacionadas a los alcances del propio diseño de la BDs, donde el modelo no permita representar las necesidades de la organización, instalación costosa, se necesita de personal especializado, etc.

- c. Investiga cuáles serían las **responsabilidades** de una **DBA** y las de un **diseñador de bases de datos**.

a) Las responsabilidades de un **DBA** son:

- Autorizar el acceso a la base de datos.
- Coordinar y monitorear el uso de la base de datos.
- Adquirir recursos de software y hardware según sea necesario.

- Asegurar la confiabilidad de la base de datos.
- Proteger la seguridad de la base de datos.

b) Las responsabilidades de un **diseñador de bases de datos** son:

- Identificar los datos que se almacenarán en la base de datos.
- Elegir las estructuras apropiadas para representar y almacenar los datos.
- Crear un diseño que cumpla con los requisitos de todos los posibles usuarios de la base de datos

d. Investiga cuáles serían los distintos tipos de usuarios finales de una base de datos, indica las principales

Las categorías de usuarios finales de una base de datos son:

- Usuario casual: acceden ocasionalmente a la base de datos con la necesidad de información diversa.
- Usuario paramétrico: consultan y actualizan constantemente la base de datos con funciones ya programadas y probadas.
- Usuario sofisticado: conocen bien el SMBD para implementar aplicaciones que cumplan requisitos complejos.
- Usuario independiente: utiliza paquetes de software específicos fáciles de usar para mantener una base de datos personal.

e. Explica las diferencias entre la **independencia de datos física y lógica**. ¿Cuál es más difícil de lograr y por qué?

La independencia de datos física es cuando los cambios en la organización física de la base de datos no afectan al mundo exterior (es decir, los programas o los usuarios directos), mientras que la independencia de datos lógica es cuando los usuarios no se ven afectados por los cambios a nivel lógico (ya sean cambios en el esquema conceptual o cambios en los esquemas externos). La independencia lógica es más difícil de lograr ya que los programas de aplicación son muy independientes de la estructura lógica de los datos a los que accede (se han creado a partir de ella).

f. ¿Qué es un **diccionario de datos** y por qué es importante para el SMBD?

Un **diccionario de datos** es un conjunto de definiciones de otros objetos del sistema. Es importante para el SMBD ya que proporciona información acerca de:

- Estructura lógica y física de la base de datos.
- Definición de todos los objetos de la base de datos.
- Espacio asignado y utilizado por los objetos.
- Valores por omisión en las columnas.
- Información acerca de las restricciones de integridad.
- Privilegios y roles otorgados a los usuarios.

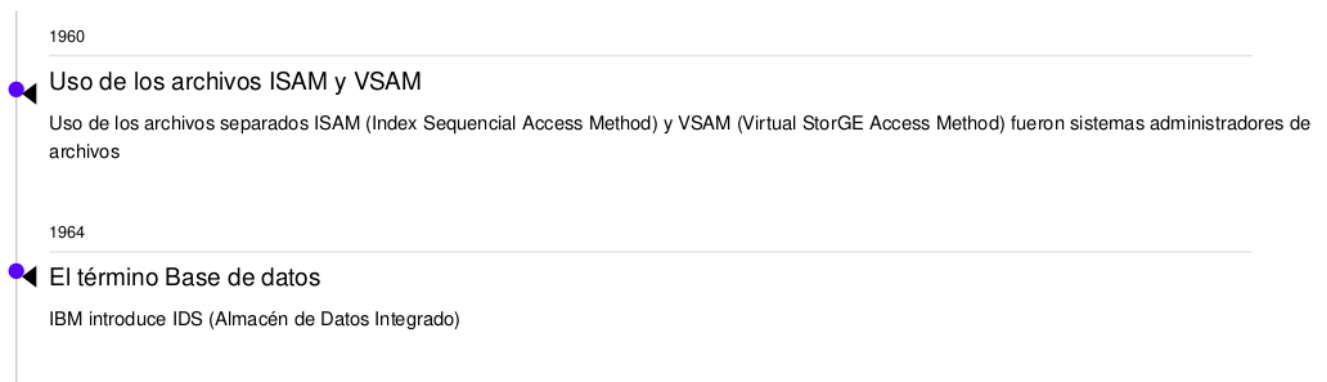
- Estadísticas de utilización, frecuencias de consultas, de transacciones y el número de accesos a diferentes porciones de la base de datos.
- g. Indica las principales características de los modelos de datos más representativos. ¿Cuáles serían las diferencias entre los modelos relacional, orientado a objetos, semiestructurado y objeto-relacional?

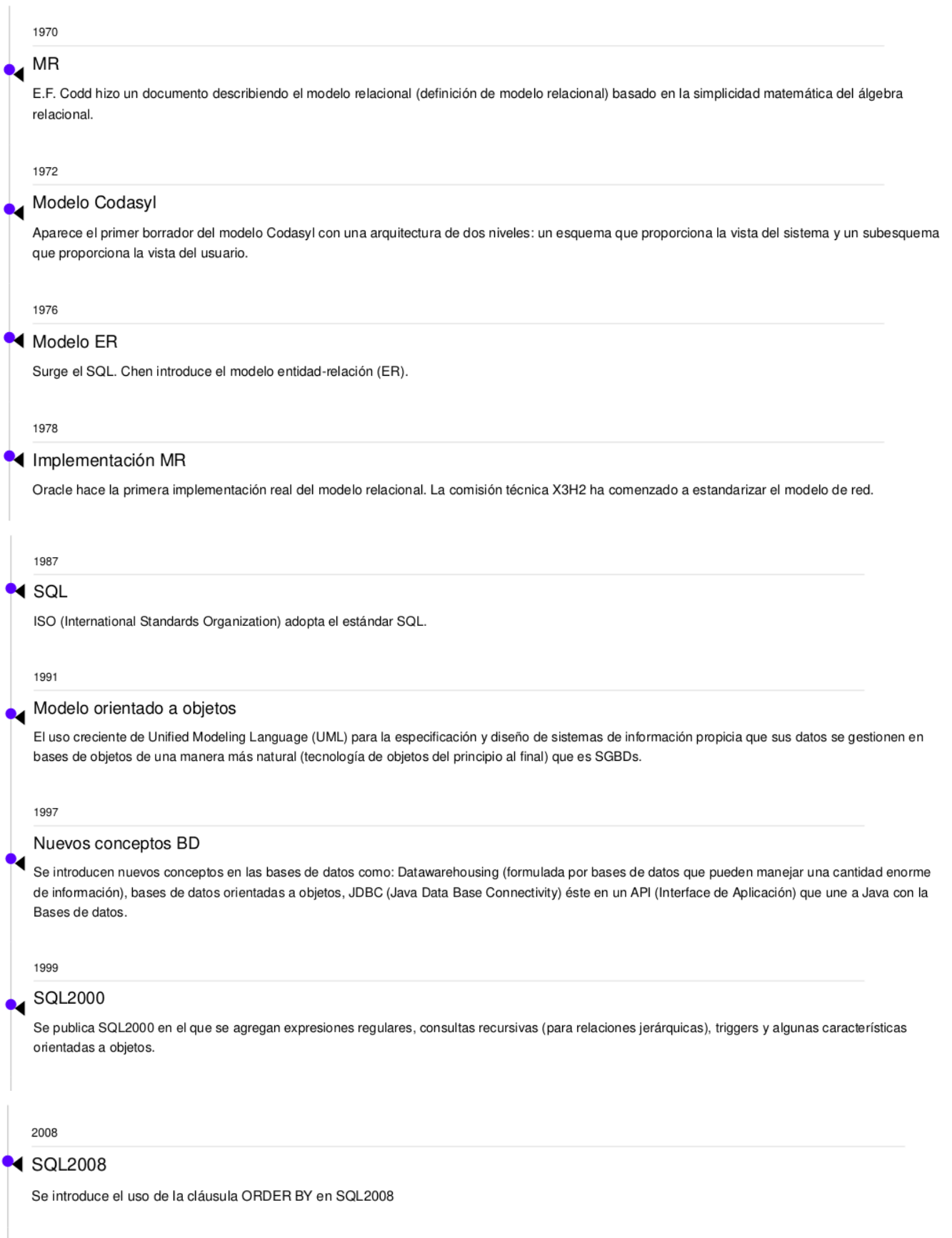
- **Modelo relacional.** Es basado en tablas. Todo el procesamiento se realiza sobre tablas y el resultado son tablas.
- **Modelo orientado a objetos.** Los datos se modelan como objetos con estado y comportamiento.
- **Modelo semiestructurado.** Representación de los datos menos rígida. Colección de nodos con su propia descripción de los datos.
- **Modelo objeto-relacional.** Aprovecha características de datos en tablas y objetos.

La diferencia entre los modelos está en cómo representan los datos, los modelos relacional y orientado a objetos tienen una estructura rígida mientras que el modelo semiestructurado es el más flexible.

- h. Elabora una **línea de tiempo**, en dónde indiques **los principales hitos** en el desarrollo de las BDs.

## *Principales hitos en el desarrollo de las Bases de Datos*





i. Indica las responsabilidades que tiene un **Sistema Manejador de Bases de Datos** y para

cada responsabilidad, explica los problemas que surgirían si dicha responsabilidad no se cumpliera.

- **Construir.** Los problemas serían que habría una inconsistencia en los datos, no habría homogeneidad en éstos ni habría un orden o una estructura en los datos. Es decir, nuestros datos estarían hechos un caos.
- **Definir.** El problema sería que no podríamos guardar los datos, por lo que habría una pérdida parcial o total de éstos.
- **Manipular.** El problema sería que no se podrían recuperar los datos perdidos, ni tampoco se podría consultar o modificar los datos.
- **Compartir.** El problema sería que más de un usuario no podría manipular la base de datos (nuestro objetivo es que varias personas puedan, de acuerdo a sus necesidades, manipular la base, cuidando en no caer en una inconsistencia de concurrencia).

j. Supón que un banco pequeño desea almacenar su información en una base de datos y le gustaría comprar el SDB que tenga la menor cantidad de características posibles. Está interesado en ejecutar la aplicación en una sola computadora personal y no se planea compartir la información con nadie. Para cada una de las siguientes características explica por qué se debería o no incluir en el SDB que desea comprar (suponiendo que se pueden comprar por separado:) **seguridad, control de concurrencia, recuperación en caso de fallas, lenguaje de consulta, mecanismo de vistas, manejo de transacciones.**

- **Seguridad.** Debemos añadirla porque, sin importar el tamaño de nuestra base de datos, nuestros datos deben estar siempre protegidos.
- **Control de concurrencia.** No es necesario incluirla. Como se trata de un banco pequeño, el número de personas que manipulen la base de datos no será grande, por lo que podemos controlar esto de forma "manual.<sup>en</sup> caso de que se llegue a presentar un problema de concurrencia (lo cual es poco probable por tratarse de pocos usuarios).
- **Recuperación en caso de fallas.** Debemos incluirla porque siempre vamos a estar expuestos a pérdidas de datos, sean intencionales o no, y en esos casos tenemos que poder recuperar los datos que necesitamos.
- **Lenguaje de consulta.** Debemos incluirlo porque siempre vamos a tener la necesidad de consultar los datos en la base y, sin este lenguaje, no podríamos manipular los datos que están en ella.
- **Mecanismo de vistas.** No es necesario incluirlo ya que no se planea compartir la información con nadie.
- **Manejo de transacciones.** Debemos incluirlo porque, a pesar de ser un banco pequeño, sin esto no podemos garantizar la integridad de nuestros datos.

## 2. Investigación.

a) ¿Qué es la **Calidad de Datos** y cómo se relaciona con las bases de datos?

Calidad de datos se refiere a las técnicas y procesos utilizados para asegurar que un dato es adecuado para su uso en operaciones y toma de decisiones. Se relaciona con las bases de datos pues debido a la gran cantidad de información que tienen se debe asegurar que todos los datos sirven su propósito ya que de no hacerlo podría tenerse información incorrecta o poner en riesgo la seguridad de la base de datos.

b) ¿Qué son las bases de datos **NoSQL**? indica el modelo de datos utilizado y algunos proveedores.

Los sistemas de bases de datos relacionales llevan décadas siendo el modelo informático más empleado del mundo para almacenar y recuperar la información.

Pero este sistema ha ido cambiando en los últimos años, surgió una contracorriente con el propósito de defender las bases de datos que no utilizaban el lenguaje SQL, denominado como el movimiento **NoSQL**.

Las bases de datos **NoSQL** son un modelo no relacional, los defensores de las bases de datos NoSQL argumentan que pueden evitar la complejidad innecesaria, proporcionan en general, un alto rendimiento, pueden manejar una gran cantidad de datos, soportar lenguajes de consulta de tipo SQL y hardware de bajo costo, entre otros.

Y estos son algunos de sus proveedores.

- Hypertable
- Cassandra
- MongoDB
- DynamoDB
- HBase
- Redis

c) ¿Qué es un **Almacén de datos**? Indica las diferencias entre éstos y una base de datos. Un Almacén de Datos (Data Warehouse) es una recopilación de datos variada en el tiempo, orientada por temas, integrada, no volátil, en apoyo de las decisiones de la gerencia. Los almacenes de datos proporcionan acceso a datos para análisis complejos, descubrimiento de conocimiento y toma de decisiones. Admiten demandas de alto rendimiento en los datos e información de una organización.

Diferencias entre un almacén de datos y una base de datos:

Características	Almacén de datos	Base de datos
<b>Cargas de trabajo admitidas</b>	Análisis, generación de informes, big data	Procesamiento de transacciones
<b>Origen de datos</b>	Datos recopilados y normalizados desde muchos orígenes	Datos registrados desde un único origen
<b>Registro de datos</b>	Operaciones de escritura masivas normalmente en un cronograma en lotes predeterminado	Optimizado para operaciones de escritura continua que datos nuevos se encuentran disponibles para maximizar el procesamiento de transacciones
<b>Normalización de datos</b>	Esquemas no normalizados	Esquemas estáticos con alto nivel de normalización
<b>Almacenamiento de datos</b>	Optimizado para acceso simple y desempeño de consultas de alta velocidad con almacenamiento de columnas	Optimizado para operaciones de escritura de alto procesamiento a un único bloque físico orientado a filas
<b>Acceso a los datos</b>	Optimizado para minimizar la E/S y maximizar el procesamiento de datos	Grandes volúmenes de pequeñas operaciones de lectura

## Referencias

- Elmasri, R. and Navathe, S. B. **Fundamentals of Database Systems**. Addison-Wesley Publishing Company, Sexta edición, 2011.
- Redman, T. C. **Data Driven: Profiting from Your Most important Business Asset**. Harvard Business Press, 2008.
- Anodrea Rodríguez M. **Introducción a Bases de Datos**. Universidad de Concepción, Chile. 2007.
- URL: <http://www.ptolomeo.unam.mx:8080/xmlui/bitstream/handle/132.248.52.100/219/A6.pdf?sequence=6>
- Avilés Rosas, G. Conceptos básicos, Fundamentos de Bases de Datos, Nota de clase.
- ¿Qué es un almacén de datos? Disponible en: <https://aws.amazon.com/es/data-warehouse/> [Acceso el 22 de agosto de 2019]