

# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

## FUNDAMENTOS DE BASES DE DATOS

Cruz Hernández Emmanuel  
Tarea 1. Conceptos Básicos.

### 1. Conceptos generales:

a) Indica las diferencias (emplea una tabla para contrastarlas) en cuanto a cómo se representan los datos entre un sistema de procesamiento de archivos y una base de datos relacional.

<b>SISTEMA DE ARCHIVOS</b>	<b>SISTEMA DE BASES DE DATOS</b>
El manejo de archivos es en texto plano	La información necesita un identificador
Se usan tipos de búsqueda secuencial, aleatorio, indexado	Las bases de datos son programas complejos y requieren mucho espacio en disco duro.
Las actualizaciones se simulan dando de baja el registro y dando de alta el nuevo con la modificación	Las actualizaciones son directas sobre el dato que se desea modificar
Tiene redundancia no controlada	Existe redundancia controlada
Tienen errores de referencia	Hay integridad de referencia
Los sistemas de archivos se organizan en carpetas	Sus componentes se organizan en archivos, registros y campos.
Están escritos para una aplicación específica	Las bases de datos están escritas para ser generales y ser útiles en muchas aplicaciones

b) ¿Qué ventajas y desventajas encuentras al trabajar con una base de datos?

<b>VENTAJAS</b>	<b>DESVENTAJAS</b>
Dependencia de datos	Se necesita personal especializado para usar bases de datos
Mejora en la consistencia	Costos considerables en la instalación y mantenimiento de las bases de datos
Control de redundancia	Costos considerables de conversión.
Mejora en compartir datos	Conflicto organizacional
Asegurar la calidad de los datos	Manejo de información complicado

c) Explica las diferencias entre independencia de datos física y lógica. ¿Cuál es más difícil de lograr? Justifica tu respuesta.

El nivel físico es el nivel más bajo de abstracción que describe cómo se almacenan realmente los datos. Se describen en detalle las estructuras de datos complejas de bajo nivel.

En cambio, el nivel lógico es el siguiente nivel más alto de abstracción que describe qué datos se almacenan en la base de datos y qué relaciones existen entre ellos. Aunque la implementación de estructuras simples en el nivel lógico puede involucrar estructuras complejas del nivel físico, los usuarios del nivel lógico no necesitan preocuparse de esta complejidad. Los administradores de bases de datos, que deben decidir la información que se mantiene en la base de datos, usan el nivel lógico de abstracción. Por esta última razón, es que el nivel lógico es más difícil de lograr, además mantiene la relación de datos entre el nivel físico y el nivel de vistas, que es el nivel más alto de abstracción que describe sólo parte de la base de datos completa.

**d) ¿Qué es el diccionario de datos y por qué es importante?**

Es un conjunto especial de tablas, también denominado directorio de datos. Un diccionario de datos contiene metadatos, que son datos acerca de los datos. El esquema de una tabla es un ejemplo de metadatos. Un sistema de base de datos consulta el diccionario de datos antes de leer o modificar los datos reales.

**e) Indica las principales características de los siguientes modelos de bases de datos: jerárquico, de red, orientado a objetos.**

El *modelo de bases de datos jerárquico* se remonta a la década de 1960. Information Management System (IMS, sistema de gestión de información), uno de los dos primeros grandes DBMS, lo desarrollaron North American Aviation e IBM para el proyecto de alunizaje Apollo. El modelo de bases de datos jerárquico no es tan flexible o fácil de comprender como el modelo relacional. Una de sus características principales es el uso del árbol como estructura de datos básica.

En el modelo de bases de datos jerárquico, a cada tipo de segmento se le asigna un código tipo único, de acuerdo con su posición en el diagrama de estructura de árbol. Los códigos tipo están determinados por el recorrido preordenado de la estructura del árbol, de modo que la raíz tiene código tipo 1, su hijo más a la izquierda tiene código tipo 2, el hijo más a la izquierda de dicho hijo tiene código tipo 3, etc.

El *modelo de bases de datos de red*, uno de los dos modelos de bases de datos más antiguos, se usó a principios de la década de 1960 en uno de los primeros sistemas de gestión de base de datos, Integrated Data Store, creado por Charles Bachman en General Electric.

La característica más sobresaliente de este modelo es que usa una representación similar para las relaciones muchos a muchos. Cuando dos tipos registro tienen una relación muchos a muchos se crea un nuevo tipo registro, llamado registro intersección o un registro vínculo, que consiste sólo de las claves de los registros asociados, más cualquier información de intersección, es decir, cualquier atributo descriptivo cuyos valores dependen de la asociación.

El *modelo de bases de datos orientado a objetos* se puede observar como una extensión del modelo E/R con las nociones de encapsulación, métodos e identidad de objeto. El modelo de datos orientado a objetos es una adaptación para los sistemas de bases de datos del paradigma de la programación orientada a objetos. Se basa en el concepto de encapsular los datos en un objeto y el código que opera sobre ellos. Una de las principales características del modelo de bases de datos orientado a objetos es que se expresa cada atributo de las entidades como una variable y un par de mensajes del objeto correspondiente. La variable se utiliza para guardar el valor del atributo, uno de los mensajes se utiliza para leer el valor del atributo y el otro mensaje se utiliza para actualizar ese valor.

**f) Elabora una línea de tiempo, en dónde indiques los principales hitos en el desarrollo de las bases de datos.**

Década de 1950:

- \* Se crean cintas magnéticas para el almacenamiento de datos. Las tareas de procesamiento de datos tales como las nóminas fueron automatizadas, con los datos almacenados en cintas.

- \* Los paquetes de tarjetas perforadas sólo se podían leer secuencialmente, y los tamaños de datos eran mucho mayores que la memoria principal; así, los programas de procesamiento de datos tenían que procesar los datos según un determinado orden, leyendo y mezclando datos de cintas y paquetes de tarjetas perforadas.

#### Década de 1960:

- \* Se introduce por primera vez el término *bases de datos* en un simposio en 1963.

- \* Se hace uso de los discos fijos a finales de la década de 1960, lo que cambió en gran medida el escenario del procesamiento de datos, ya que los discos fijos permitieron el acceso directo a los datos.

- \* Los datos se liberaron del uso de la secuencialidad.

- \* Con los discos pudieron desarrollarse las bases de datos de red y jerárquicas, que permitieron que las estructuras de datos tales como listas y árboles pudieran almacenarse en disco.

- \* Se presenta la alianza de IBM y American Airlines para desarrollar SABRE, que es un sistema operativo creado para manejar transacciones e información sobre los pasajeros de la compañía American Airlines.

- \* En 1969 se propone separar el nivel lógico del físico.

#### Década de 1970:

- \* El modelo relacional se propone por primera vez por E. F. Codd en 1970, en un ensayo titulado “A Relational Model of Data for Large Shared Data Banks”.

- \* Lawrence J. Ellison desarrolló lo que actualmente se conoce como Oracle Corporation.

- \* Se desarrolló SQL, un lenguaje de consultas que permite efectuar consultas para recuperar información.

#### Década de 1980:

- \* Comienzan a comercializarse los sistemas relacionales.

- \* SQL comenzó a ser estándar en la industria.

- \* A los principios de la década de 1980 las bases de datos relacionales llegaron a competir con los sistemas de bases de datos jerárquicas y de red incluso en el área de rendimiento.

- \* Se presenta una gran investigación en las bases de datos paralelas y distribuidas, así como del trabajo inicial en las bases de datos orientadas a objetos.

#### Década de 1990:

- \* La ayuda a la toma de decisiones y las consultas reemergieron como una importante área de aplicación para las bases de datos.

- \* Muchos vendedores de bases de datos introdujeron productos de bases de datos paralelas en este periodo, así como también comenzaron ofrecer bases de datos relacionales orientadas a objetos.

- \* Se desarrollaron nuevas herramientas como Excel y Acces.

- \* A finales de 1990 se presenta el crecimiento explosivo de World Wide Web.

- \* Los sistemas de bases de datos tienen ahora soporte para tasas de transacciones muy altas, así como muy alta fiabilidad y disponibilidad de 24 horas al día y 7 días a la semana, que significa que no hay tiempos de inactividad debidos a actividades de mantenimiento planificadas.

- \* Los sistemas de bases de datos también tuvieron interfaces Web a los datos.

#### Siglo XXI:

- \* Las tres grandes compañías dominantes del mercado de las bases de datos son IBM, Microsoft y Oracle.

\* Google se convierte en la compañía con más generación de información.

**g) Explica cuál(es) de los siguientes conceptos juegan un papel importante en la representación de los datos acerca de un problema real en una base de datos y por qué: lenguaje para definición de datos, lenguaje para manipulación de datos, el modelo de datos, manejo de transacciones.**

El lenguaje para definición de datos si juegan un papel importante para la representación de datos, ya que a través de este, se llevan a cabo las tareas de definición de las estructuras en las que se almacenarán los datos, además permite hacer tareas usando las funciones para poder consultar los datos.

El lenguaje de manipulación de datos no tiene un papel tan importante en la representación, ya que este se encarga de consultar o modificar datos, pero esto es independiente de la representación de los mismos.

El modelo de datos si es importante para la representación de los datos, porque nos permite poner restricciones de integridad sobre los datos, además de que nos da una base sólida para saber el tipo de datos que hay en la base, así como saber la forma en que se relacionan.

El manejo de transacciones no es algo fundamental para la representación de la base de datos, que no afecta directamente con la estructura o manera en que representan los datos.

**h) Supón que una pequeña compañía desea almacenar su información en una base de datos. Desea comprar la que tenga la menor cantidad de características posibles, se desea 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 la base de datos que se 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.**

No se necesita seguridad porque no se planea compartir la información con nadie, entonces no hay problemas respecto a la seguridad de datos.

Como la base de datos se va a ejecutar en una sola computadora personal, entonces no es necesario controlar la ejecución de transacciones en paralelo, ni controlar el acceso a información compartida, por lo que no es necesario el control de concurrencia.

La recuperación en caso de fallas si es importante comprarla. Independientemente de que la aplicación sea para una computadora es necesario prevenir algún problema que no afecte los datos almacenados, como pérdidas totales o parciales de información.

El lenguaje de consulta depende del usuario y el fin de la base de datos. Si se pretende hacer consultas en la base de datos y sistemas de información, entonces si es necesario comprar el lenguaje de consultas, de lo contrario, no lo es.

Como en el caso anterior, si se pretende hacer consultas presentadas como una tabla a partir de un conjunto de tablas, entonces es buena idea comprar un mecanismo de vistas.

En una base de datos es importante conservar la consistencia de datos. Cuando se hace una transacción se hacen transformaciones de los estados, por lo que es necesario comprar un manejo de transacciones.

## **2. Investigación**

**a) ¿Qué se es la Ciencia de Datos y Big Data? ¿Cómo se relacionan con las bases de datos?**

La ciencia de datos es un proceso para descubrir información en grandes cantidades de datos, esto involucra el conocimiento de uno o más dominios, técnicas científicas como hipótesis y pruebas,

resultados más confiables, matemáticas y estadísticas, aprendizaje automatizado, inteligencia artificial o algoritmos de descubrimiento de conocimiento. Estas ramas o campos, son muy útiles para lograr el objetivo de la ciencia de datos, que es descubrir información.

Big Data son datos que contienen una mayor variedad y que se presentan en volúmenes crecientes y a una velocidad considerable. Esta va más allá de las bases de datos relacionales, además permiten resultados que no están disponibles con los enfoques ya conocidos o que llevarían mucho más tiempo.

Se relacionan con las bases de datos porque se genera mucha información que debe ser organizada de alguna manera con ayuda de una base de datos. Ya que, además de proteger los datos, se debe tener un enfoque de consistencia a la información y no permitir redundancia.

**b) Especifica las características más importantes de las bases de datos NoSQL, indica el modelo de datos que utilizan y principales proveedores.**

Las bases de datos NoSQL manejan grandes cantidades de información. Uno de los grandes problemas de las bases de datos relacionales es que se deben comprar equipos con suficiente capacidad para mantener la base de datos. En cambio, con las bases de datos NoSQL, basta con mover la información a la nube en entornos virtualizados, por lo que no es necesario invertir tanto en hardware.

Las bases de datos tienen la ventaja de que son creadas para requerir la reparación automática mínima, distribución de datos y modelos de datos simples. Estas bases de datos, también suelen mantener servidores genéricos baratos para administrar los datos y el volumen de transacciones.

Otra de las características principales es que no dependen de algún modelo específico para usarlas, lo que permite flexibilidad para almacenar información.

Algunas de las empresas que usan bases de datos NoSQL son Facebook, Amazon, Twitter y Google.

## **FUENTES CONSULTADAS**

Silberschatz Abraham, “Fundamentos de Bases de Datos”, McGrawHill, Cuarta Edición, Madrid, 787 páginas.

Catherine M. Ricardo, “Bases de Datos”, McGrawHill, Primera edición, México, 665 páginas.

<https://www.oracle.com/mx/big-data/guide/what-is-big-data.html>

<https://www.ibm.com/analytics/es/es/technology/data-science/>

[http://moodle2.unid.edu.mx/dts\\_cursos\\_md1/lic/TI/FB/AM/01/Historia\\_de\\_las\\_bases\\_de\\_datos.pdf](http://moodle2.unid.edu.mx/dts_cursos_md1/lic/TI/FB/AM/01/Historia_de_las_bases_de_datos.pdf)