



Ingeniería de Datos

Introducción a las bases de datos y su ciclo de vida

Agenda

- Introducción.
- Características de una Base de Datos.
- Tipos de Base de Datos.
- Administrador de Base de Datos.
- Fases en el Diseño de una Base de Datos.

Introducción

- ¿Cómo los negocios en las empresas generan datos?
- ¿Por qué requerimos almacenar los datos en una compañía?
- ¿Cómo deberían ser almacenados los datos?
- ¿Los datos son confiables en su calidad?
- ¿Cómo consultamos los datos?
- ¿Cuáles son los problemas que se presentan al almacenar los datos sin un buen diseño?

Dato vs Información

- Los datos son las unidades básicas de la información no procesada, los cuales por si solos no nos ofrecen un significado, requieren de un análisis de su contexto para interpretar su significado.
- La información es el resultado del procesamiento de los datos, los cuales nos entregan un resultado o respuesta significativa. Esta respuesta, finalmente, nos lleva a tomar decisiones.

¿Qué es una Base de Datos?

- ... Es un conjunto de datos persistentes utilizado por los sistemas de aplicaciones de una empresa determinada.
- ... Es una colección de datos interrelacionados creados con un DBMS.
- ... Es un conjunto auto descriptivo de datos integrados.
- ... Sirve para poder almacenar datos que luego pueden ser recuperados al hacer una consulta y convertirse en “información”.

Características de una Base de Datos

- CRUD (Create / Read / Update / Delete).
- Consistencia (Los datos deberían ser los mismos en el transcurso del tiempo si no ha tenido modificaciones).
- Validez (Datos validados).
- Corrección fácil de los errores.
- Velocidad.
- ACID (Atomicity / Consistency / Isolation / Durability).
- Persistencia & Copias de Seguridad.

¿Qué es una Base de Datos?

- Portabilidad (Descargar datos relevantes de forma sencilla).
- Seguridad.

(<https://www.privacyrights.org/data-breach>)

August 9, 2013 Auburn University - School of Forestry and Wildlife Sciences
Auburn, Alabama

EDU DISC

Unknown

Spreadsheets with donor and alumni information were accidentally uploaded to a publicly accessible server after an administrative error. The error was discovered on June 19 and Auburn's IT office removed the information. Names, Social Security numbers, maiden names, mailing addresses, first year at Auburn, graduation year, alumni status, email addresses, and phone numbers were exposed.

Information

Source:

Databreaches.net

records from this breach used in our total: 0

Manejadores de Base de Datos (DBMS)

- Es básicamente un sistema para almacenar archivos en un computador.
- Su propósito es reguardar la información de dichos archivos y hacer que esté disponible cuando se la solicite.
- Para la mayoría de los DBMS corporativos, la información estará “integrada” y “compartida”.
- Sistema de Gestión de Base de Datos (SGDB) permite el almacenamiento, modificación y extracción de la información en una base de datos.

Manejadores de Base de Datos (DBMS)

- Sus funciones son:
 - Crear, organizar y manipular la Base de Datos.
 - Definición de Datos (DDL).
 - Manipulación de Datos (DML).
 - Mantener Sub Sistemas de Seguridad, Integridad, Control de Conurrencia y Recuperación.
 - Registrar las estadísticas de uso.
 - Administrar el diccionario de datos.
 - Optimizar el desempeño.

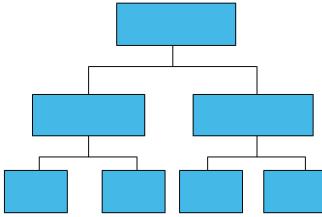
Tipos de Base de Datos

- Archivos Planos.
- Jerárquicas.
- Relacionales.
- Hojas de Cálculo.
- Redes.
- Objetos.
- Orientada a Documentos.
- Dimensionales.



Archivos Planos

- Los archivos planos son sencillamente archivos que contienen texto.
- No ayudan a buscar datos y no aportan funciones de cálculos matemáticos.
- No soportan características especiales como mecanismos de bloqueos, seguridad, backups, etc.
- Útil cuando la información a almacenar es relativamente sencilla y raramente cambia.



Jerárquicas

- Los datos jerárquicos incluyen valores que se organizan de forma natural en una estructura de árbol. Ejemplos: Estructura de un Libro, el árbol genealógico de una familia, Regedit de Windows, etc.
- Funcionan bien si:
 - Los datos tienen una naturaleza jerárquica.
 - Realiza operaciones que sacan provecho a la estructura jerárquica.
- XML es la estructura más popular en este tipo de base de datos.

Relacionales

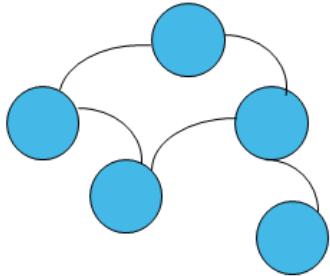
- Contienen tablas que a su vez contienen filas y columnas. Cada fila contiene datos relacionados de una entidad particular (persona, vehículo, etc). Cada columna representa una parte de los datos de esa entidad (nombre, apellido, etc).
- Las relaciones se establecen entre atributos comunes de las tablas.
- En el mercado se les conoce como RDBMS y lo más importantes son: Oracle, MSSQL, DB2, MySQL, Teradata e Informix.

Hojas de Cálculo

- Tienen filas y columnas de datos.
- Admite herramientas de análisis relativamente sofisticadas como las funciones estadísticas y la búsqueda de soluciones iteradas.
- No admiten consultas complejas ni comprueban de forma automática la integridad de los datos por que es fácil introducir valores incorrectos o inconsistentes.

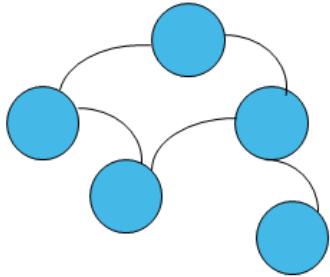
Hojas de Cálculo

- No permite actualización de grandes cantidades de datos.
- No funciona bien cuando existe relaciones complejas entre los valores de distintas hojas de trabajo.



Redes

- Una red contiene una colección de nodos que están conectados mediante enlaces y los enlaces representan todo tipo de cosas como por ejemplo: líneas telefónicas, calles, circuitos digitales, organigramas, etc.
- Útiles cuando se tienen que realizar operaciones de red en los datos como encontrar la ruta más corta o calcular flujos de red.



Redes

- Algunos sistemas operativos suelen implementar un modelo a redes para dar soporte a sus directorios, archivos y soft links.
- Neo4J es un software de base de datos que implementa un tipo de base de datos a redes.

Objetos

- Se utilizan abstracciones de programación llamados objetos para representar elementos como clientes, productos, etc. Una base de datos de objetos gestiona objetos.
- Proporciona herramientas para guardar los cambios en un objeto en la base de datos y asimismo una sintaxis de consulta para recuperar objetos desde la misma.
- Conocidas en el mercado como OODBMS y son

Objetos

- Funciona bien si la arquitectura de programación se ve favorecido por la utilización de objetos y no tiene que realizar consultas complejas.
- Solo permite ser consultado por la aplicación por lo cual no existe el concepto de procedimientos almacenados.
- Esquema híbrido entre Relacional & Objeto es: Base de Datos Objeto Relacional (ORDBMS).

Orientada a Documentos

- Una base de datos orientada a documentos está constituida por un conjunto de programas que almacenan, recuperan y gestionan datos de documentos o datos de algún modo estructurados.
- Este tipo de bases de datos constituyen una de las principales sub categorías dentro de las denominadas bases de datos no SQL.

Orientada a Documentos

- Los documentos dentro de una base de datos orientada a documentos son similar, de algún modo, a registros, tuplas o filas en una base de datos relacional pero menos rígidos.
- Por ejemplo un documento puede ser:

```
{  
    Nombre:"Pepe",  
    Dirección:"Plaza Mayor 5",  
}
```

Orientada a Documentos

- Los documentos contienen alguna información similar y otra diferente entre ellos. Al contrario que una base de datos relacional en la que todos los registros deben tener los mismos atributos.
- Un documentos representa un registro y una colección de documentos representa una tabla comparándolo con la base de datos relacional.
- Solo se puede realizar transacciones a nivel de un documento.

Orientada a Documentos

- Muy útiles para aplicaciones e-commerce y no sería útil para aplicaciones bancarias.
- MongoDB es una de las base de datos más conocidas como base de datos orientada a documentos.

Dimensionales

- Llamado también multidimensional.
- Representa distintos aspectos de los datos en dimensiones en lugar de tablas separadas en una base de datos relacional.
- Se puede pensar en ellos como una caja rectangular multidimensional en la que cada dimensión presenta alguna faceta importante de los datos.
- Son útiles para buscar patrones en los datos.

Administrador de Base de Datos

Las responsabilidades del Administrador de Base de Datos (DBA) son:

- Evaluación del Hardware donde residirá la base de datos.
- Instalación del software de base de datos.
- Planificación de una estrategia de seguridad.
- Creación y migración de la base de datos.
- Respaldos de la base de datos.
- Diseño en la conectividad de las aplicaciones y de los usuarios a la base de datos.
- Implementación del diseño físico de la base de datos.
- Restauración de la base de datos ante un incidente.
- Monitoreo del desempeño de la base de datos.

Fases en el diseño de una Base de Datos

-
- Modelo Conceptual
 - Recolección
 - Identificación de Requerimientos.
 - Determinar el conjunto mínimo de elementos de datos requerido para dar soporte al sistema.
 - Refinamiento.
 - Identificación de las entidades y sus atributos.
 - Descripción y documentación de entidades (cardinalidad, frecuencia de consultas).
 - Identificar las relaciones simples entre elementos de datos.
 - Normalizar.

Fases en el diseño de una Base de Datos

- Modelo Lógico.
 - Determinar las relaciones entre entidades.
 - Asegurar que los requerimientos del sistema puedan ser soportados por la base de datos.
 - Depende del tipo de DBMS a emplear.
- Modelo Físico
 - Determinar la representación física de las entidades y sus atributos (tablas) a ser implementados en un DBMS.