

# Bases de datos avanzadas

## Contents

Bases de datos avanzadas .....	1
Evaluaciones .....	1
Temas .....	1
Primer Parcial .....	2
Datos sensibles .....	2
DBMS .....	2
Dato .....	2
Registro .....	2
Entidad (o tabla) .....	3
Información .....	3
Objetos .....	3
ORM .....	3
DBMS (SMBD en es.) .....	4
Funciones .....	4
Términos .....	4
Proyecto .....	4
Respaldo .....	4
Replicación .....	5
Concurrencia .....	5
Transacciones .....	5
Componentes .....	5

—

## Evaluaciones

1. Parciales | 1.1 con el examen ... 80 % | 50 % ? 1.2. Otra cosa ... 20 % |
3. Proyecto final (no acumulativo) ... 25 % | 25 % – 2 sem antes del examen final  
3.1 Desarrollar un dashboard de información sobre una base de datos. En MySQL o Postgres. 3.2 No nos va a dejar que nos lo imaginemos, el nos va a dar un ejemplo exacto de renta de videos. 3.3 Se califica: 3.3.1 Los datos pueden ser de donde sea pero tienen que ser masivos 3.3.2 Identificar los queries importantes que se deben mostrar al usuario:
  - No es un sistema que administre datos, solo es una hoja con graficas que muestra el comportamiento del negocio
  - El chiste es hacer que los datos le hablen al dueño de la empresa, hacer visualizaciones agradables y eso
  - Quien compra mas, edades, donde, cómo no el otro lado.3.4 Pasos 3.4.1 Determinar una necesidad, o algo que a nosotros nos guste mucho 3.4.2 Proponer la respuesta, como el intentar indicar las esquinas donde mas violencia hay y de que tipo. 3.4.3 Pensar en hacerlo lo más amigable a alguien que lo que quiere es ver datos para tomar decisiones. 3.5 Siempre tener una copia local, no vaya a ser que falla la nube. 3.6 Lenguaje el que queramos. 3.7 Seguridad de las personas o salud
4. Examen final ... 25 % | 25 %

Bases de datos estructuradas con SQL, el proyecto en equipo de 2 a 3.

## Temas

1. Queries, queries con joins, triggers, normalización.
2. Replicación, compactación.
3. Bases de datos distribuidas, orientadas a objetos
4. Presentaremos temas relacionados a: Data mining, data warehouse, big data

## Primer Parcial

### Datos sensibles

Para poder datos sensibles se necesita autorización expresa de la persona y cifrar la base de datos que los contenga.

El genero no es un dato sensible, pero si personal.

1. Datos personales: Nos permiten identificar a una persona, se pueden usar siempre y cuando le digamos al usuario cómo vamos a usar sus datos.
2. Datos patrimonial: Datos personales como cuentas de banco, tarjetas, números, asociaciones, autos, donde se involucra dinero. Esos datos requieren un permiso firmado por autografía de la persona.
3. Datos sensibles: Permiten determinar preferencias de la persona, esto puede provocar discriminación del individuo, por lo que se requiere tener cuidado.

En kaggle podemos sacar datos

### DBMS

Es un programa que permite llegar a los datos, manejarlos.

Lo que se suele conocer como base de datos en realidad se refiere al sistema manejador de bases de datos (DBMS por sus siglas en inglés). Los datos se almacenan, en disco en archivos binarios. El sistema manejador es el que se encarga de manejar los datos, así nosotros podemos manejarlos de forma sencilla.

El DBMS tiene distintos componentes en su arquitectura:

- BD: El archivo donde se guardan los datos
- Manejador de BD: Es el programa que se ocupa para administrar los datos, ejemplo Maria, MySQL, etc. Es un servicio, un demonio.
- Administrador del manejador de BD: Es el programa con el que nos comunicamos con el servicio, es el cliente.

Daemon, que es un servicio, viene del griego. Recordemos que a los muertos se les ponían ponedas en los ojos para Caronte, quien no tenía para pagar se quedaba en el lado de los vivos, pero muerto, por lo que los vivos no lo ven pero tampoco está muerto. Por eso un servicio se le llama daemon, porque los vivos no lo ven

### Dato

El dato es algo que podemos corroborar con la realidad. Puede ser:

- Hecho: *hoy es X*
- Antecedente
- Fundamento

Los datos están asociados con un *nombre*, un *tipo* y un *tamaño*.

### Registro

Un registro es un conjunto de datos, con una relación entre ellos.

## Entidad (o tabla)

Donde se guardan los registros.

## Información

Conjunto de datos ordenados en cierto:

- **Formato:** Como estan organizados los datos. La manera en que se almacenan es diferente, lo que permite realizar distintas operaciones (?)
- **Contexto:**
- **Disminuye la incertidumbre:** Ayuda al que otro entienda bien. Que no hay certeza, no se entiende bien.
- **Para la toma de decisiones:** Si no hay decision entonces no es información, si se escucha y no se actua no es información.

La cultura es contexto, debido al contexto que tenemos las mujeres en occidente se casan de blanco, en cambio en asia es de rojo. Por el contexto se ve el mundo de forma totalmente diferente, con otra perspectiva.

La información es subjetiva por definición, esta rodeado de contexto. Dependiendo de la persona, el mismo hecho puede representar información y para otros no.

Las bases de datos tienen datos, no información. Podemos explotar las bases de datos para ayudar al otro a constituir en él información. Cuando un bot toma decisiones basado en datos si se constituye información, al final fue un humano quien lo decidió y quien actua por medio de un agente.

## Objetos

- **Tabla/Entidad:** Con objeto nos referimos a elementos que la base de datos sabe almacenar.
- **Vista:** Es un sql que se almacena en la base de datos y se muestran como tabla. No se puede escribir en una vista, pero si se puede usar como una tabla.

Las vistas son solo de escritura, se puede inventar campos, no está *materializado* en ningun archivo o tabla, por lo que no hay dónde escribir.

- **Stored procedures:** Código SQL almacenado en el servidor que no se muestra como tabla, sino que se ejecuta, es como una función, así que admite in y out.
- **Jobs:** Como un *chron* en Unix. Tiene un disparador (*trigger*) y un código que ejecuta. El código puede ser un *stored procedure*. Para respaldos, calculos, etc.
- **Triggers:** Otro tipo de disparadores, asociado a eventos de lectura/escritura. Se asocia a las operaciones que pueden afectar la integridad. Por lo tanto, se suele usar cuando tenemos operaciones que pueden alterar la integridad referencial de la base de datos. Tiene disparador (evento io) y código.
- **Constraints:** Reglas que podemos tener de relación entre tablas, que dejan condiciones para las operaciones con los datos que mantengan la integridad de la base de datos.
- **Formulario:** Solo lo tiene Access, puedes guardar y mostrar/preguntar al usuario que rellene bases de datos.
- **Usuarios:** Los '*usuarios*' son contraseña/usuario configurado para dar permisos especificos a quienes ingresen a la base de datos. Lo que se refiere con que no son los mismos usuarios es que no hay un usuario 1:1.

## ORM

**Historia:** Parte de las prácticas que realizan las bases de datos para conservar sus clientes. Por eso existen bases de datos con planes tipo: Hago el estándar de SQL 92 y *esto más*. A estas modificaciones se les llama dialectos, haciendo a la vez más difícil migrar entre bases de datos.

**Patrones de diseño:** Debido a los patrones de diseño se pueden llegar a formas comunes de crear software, estos son los patrones de diseño.

No conviene realizar bases de datos que empleen la tecnología específica de la base de datos, como triggers y procedures. Un *Object Relational Mapping* provee una forma en que podemos usar con la misma interfaz distintas bases de datos.

Si mi aplicación emplea un ORM podemos desarrollar aplicaciones sin necesidad de casarse con una base de datos específica.

## DBMS (SMBD en es.)

Es un conjunto de varios procedimientos/programas para que podamos recuperar ({} SELECT), describir (CREATE/ALTER) y manipular (INSERT/UPDATE/DELETE) datos que se almacenan en la base de datos.

Esto de forma que se mantenga la:

- integridad: que se mantenga válida la información y sus relaciones
- confidencialidad: que solo pueda acceder quien tiene permiso a lo que tiene permiso
- seguridad: asegurar los datos.

## Funciones

Lo mismo, describir, manipular y utilización (accessible) de los datos. Si los datos no llegan no hay chiste.

- Interacción con el fs: Antes se especificaba SO pero ahora se confía en el nativo
- Implantación de seguridad: Lo mismo de las operaciones peligrosas
- Seguridad: Verifica que los accesos a la base de datos estén realizados por los usuarios. Verificar brechas de seguridad, monitorear.

El 60% de los ataques en las empresas son de empleados

## Términos

- **Integridad referencial:** Hacer referencia a otro dato, lo que implica que el otro existe. Que las llaves foráneas/primarias si estén relacionadas.
- **Operaciones fundamentales:** Leer, insertar, modificar, borrar

## Proyecto

Mostrar controles, todo fácil de digerir, chance simular. Lo importante es extraer los datos.

## Respaldo

Tenemos un disco, hacemos una copia de respaldo, almacenamos las copias en lugares seguros, distribuidos, fáciles y rápidos de acceder.

En la nube puede ser un lugar pero hay que considerar los tiempos de subida/bajada.

Se debe ejercitar el poder respaldar, simulacros donde se pruebe la facilidad de respaldo, todo el ciclo del mismo.

El respaldo se debe hacer considerando la cantidad de información que entra, las operaciones importantes que suceden con la importación, etc. Cada que se haga algo relevante.

## Replicación

En el caso de sistemas con cantidades de eventos relevantes altas en muy poco tiempo se debe tener *replicación*, otra(s) base(s) de datos que sean espejo en tiempo real.

## Concurrencia

El SMBD se encarga de manejar la concurrencia de forma que no haya colisiones, donde se corrompen los datos desde el nivel de memoria.

## Transacciones

Una transacción en el contexto de la base de datos es un conjunto de tareas para la base de datos. O se hacen exitosamente todas las tareas o se cancela el grupo como conjunto.

## Componentes

- Tenemos el **gestor de seguridad**, que valida que los usuarios (*de acceso*) pueden entrar cuando deben de entrar
- El **gestor de consultas** que se encarga de recibir las consultas y los almacena en una cola
- El **optimizador de consultas** verifica si puede modificar los queries (sin cambiar su efecto) para hacer el query más rápido, así podemos atender más consultas en menos tiempo.
- El **planificador**, o *scheduler*, el orquestador de tareas se encarga de dar prioridad a las consultas, es decir, influir en el orden en que las consultas se ejecutan para evitar colisiones, optimizar los recursos y reducir el tiempo de ejecución.
- **Procesador de consultas/transacciones** ejecuta el query.
- **Gestor de archivos**, que se encarga de administrar los archivos, en estos tiempos no se nota tanto pero en la antigüedad optimizaba la escritura en momento adecuado.
- **Buffers**: Transfiere datos entre memoria y secundarios, los buffers son el paso intermedio, si se apaga el servidor es necesario que se almacene, puede ser parte de la RAM.
- El **Gestor de recuperación** garantiza la consistencia de la base de datos.

Los SMBD hoy en día no se almacenan en SSDs, porque son caros y no alcanzan el tamaño que si pueden almacenar discos de estado sólido. Además se puede utilizar el esquema *raid* que permite almacenar copias entre múltiples discos para siempre asegurar la información.