

PROYECTO COMPILANDO CONOCIMIENTO

PROGRAMACIÓN

Bases de Datos

Una Pequeña (Gran) Introducción

AUTORES:

Rosas Hernandez Oscar Andrés

Lopez Manriquez Angel

Índice general

I	Parte Abstracta	3
1.	Definiciones	4
1.1.	Repositorio de Datos	5
1.2.	Base de Datos	5
1.3.	Proposito para una Base de Datos	5
1.4.	Elementos de un Sistema de Base de Datos	8
1.5.	Abstracción de los Datos	9
1.6.	Usuarios de una Base de Datos	10
1.6.1.	Usuarios Finales	11
1.7.	DML vs DDL	12
1.8.	Arquitectura del Sistema Gestor	14
1.9.	Arquitecturas de las Bases de Datos	16
2.	Modelos de Datos	17
2.1.	Sistema Entidad-Relación	18
2.1.1.	Entidades	19
2.1.2.	Relaciones	20
2.1.3.	Cardinalidad	20
2.1.4.	Atributos	21
2.2.	Sistema Entidad-Relación Extendido	22
2.3.	Sistema Relacional	24
2.3.1.	Relaciones	25

3. SQL	26
3.1. SQL como Lenguaje	27
 II Parte Practica	 28
 4. Queries en SQL	 29
4.1. SELECT FROM	30
4.1.1. Distint vs All	30
4.2. Funciones de Agregación	30
4.3. WHERE	31
4.4. Comparación de Strings	31
4.5. Orden de las Tuplas	31

Parte I

Parte Abstracta

Capítulo 1

Definiciones

1.1. Repositorio de Datos

Son un conjunto de datos, donde definio a un dato como cualquier información que sea válida.

1.2. Base de Datos

Son un conjunto de datos interrelacionado definido por un modelo de datos, esto no lo tiene necesariamente un repositorio de datos. Así como progrmas que nos permitan acceder y manipular esa información.

Podemos definir de manera alterna como: Una colección de registros el cual es almacenada en una computadora de una forma sistemática (estructurada), de tal forma que un programa de computadora pueda consultarlo para responder consultas.

1.3. Proposito para una Base de Datos

Todo muy bien, pero ¿porque debería importarme un comino?¿Porqué preferir una base de datos sobre simplemente guardar los datos de manera “común”?

Antes de la aparición de los SGBD, las organizaciones normalmente almacenaban la información en Sistemas de Procesamiento de Archivos Típicos (Sistemas de Archivos).

Un sistema de archivos es un conjunto de programas que prestan servicio a los usuarios finales, donde cada programa define y maneja sus propios datos, los cuales presentan los siguientes inconvenientes:

- **Redundancia de datos e inconsistencia:** Ya que tu no vas a programar un sistema entero habra muchas maneras en que los demás programadores crearán aplicaciones y sobretodo en como van a guardar los datos.

Peor aún, ¿Qué pasa cuando tengamos un motón de archivos con casi la misma información? Es decir cuando tengamos un montón de archivos con tu mismo número de telefono, con tu misma información de contacto.

- **Problemas para acceder a la información que queríamos en primer lugar:** Supongamos que queremos acceder a los datos, digamos que tenemos un montón de registros de sobre alumnos.

¿Como haríamos para tener todos los alumnos que hayan reprobado? No hay forma fácil de hacerlo, incluso la forma mas “correcta” sería desarrollar un pequeño programa que se encargue de hacer lo que queremos.

Y esto podría servir muy bien.... Hasta que necesitemos algo más. Entonces tenemos que estar haciendo programas y programas ¡Que cansado!

- **Aislamiento de la Información:** Ya que puede estar repartida por todos lados, no se exactamente por donde tengo que empezar a buscar. Y esto se vuelve un verdadero desastre cuando intentamos modificar la información guardada.

- **Problemas con Integridad:** Los valores de los datos almacenados en la base de datos, deben satisfacer ciertos tipos de restricciones de consistencia.

Por ejemplo, el saldo de cierto tipo de cuentas bancarias no pueden ser nunca inferior a una cantidad predeterminada, digamos 4000 dolares.

Los desarrolladores deben cumplir estas restricciones en el sistema añadiendo el código correspondiente en los diversos programas de aplicación. Sin embargo, cuando se añaden nuevas restricciones, es difícil cambiar los programas para hacer que se cumplan.

- **Problemas con Atomicidad:** Si ocurriera un problema en el sistema al momento de estar modificando la información me gustaria que al volver a arrancar todo, este debería regresar a un punto de respaldo.

Como si el sistema fallara justo al hacer una transacción bancaria, me interesa que se haya realizado o no, pero que no le haya quitando dinero a una cuenta pero no se la haya dado al otro cliente.

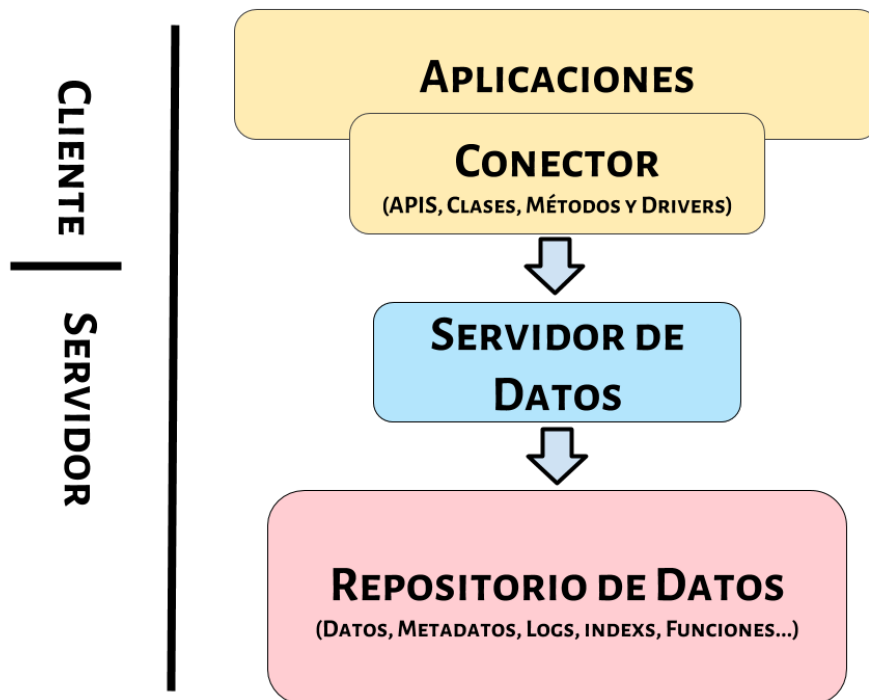
- **Errores del Acceso Concurrente:** Para aumentar el rendimiento global del sistema y obtener una respuesta más rápida, muchos sistemas permiten que varios usuarios actualicen los datos simultaneamente. En realidad hoy en día, los principales sitios de comercio electrónico en internet pueden tener millones de accesos diarios de compradores a sus datos. En tales entornos es posible la interacción de actualizaciones concurrentes y puede dar lugar a datos inconsistentes.

- **Problemas de seguridad:** No todos los usuarios de un sistema de base de datos deben tener acceso a todos los datos. Ya que los programas de aplicación se añaden al sistema de procesamiento de datos de un forma adhoc, es difícil hacer cumplir tales restricciones de seguridad.

Podemos resumir esto con que a diferencia de un sistema de archivos, una base de datos busca:

- Evitar o miniza la redundancia
- Evitar inconsistencias en los datos
- Eliminar inconsistencias en los datos
- Comunicacion con distintos repositorios de datos
- Control de concurrencia

1.4. Elementos de un Sistema de Base de Datos



- **Aplicaciones:** Es la interfaz entre la base de datos y el usuario, estas pueden ser desarrolladas por un lenguaje de alto nivel
- **Conector:** Son los componentes que permiten el enlace entre el SGBD y las interfaces desarrolladas en un lenguaje de programación, estas contienen las clases y/o funciones necesarias para llevar a cabo la comunicación entre las aplicaciones con el Sistema Gestor de Base de Datos.
- **Sistema Gestor de Base de Datos:** Son el software especializado que nos permite manipular inteligentemente nuestros datos:

Es la aplicación que permite a los usuarios definir, crear y mantener la base de datos y proporciona acceso controlado a la misma.

- Creación de Repositorios
- Creación de Cuentas de Usuarios
- Se encarga de crear archivos lógicos, físicos y objetos de la BD.
- Se encarga de administrar las transacciones, bloqueos, etc.

1.5. Abstracción de los Datos

- **Nivel Físico:**

Es el nivel base de abstracción que describe como es que la información es guardada de manera actual. Esto nos permite describir como de complejo es que son las estructuras en la realidad.

- **Nivel Lógico:**

Es el nivel que nos muestra como es que se almacena la información dentro de la base de datos y como es que se da la relaciones entre la información.

Aunque la implementación de las estructuras simples en el nivel lógico puede implicar complejas estructuras de nivel físico, el usuario de este nivel no necesita ser consciente de esta complejidad. Esto se conoce como independencia de datos físicos. Los administradores de bases de datos, que deben decidir qué información debe conservar en la base de datos, utilizan el nivel lógico de abstracción.

- **Nivel Visual:**

El nivel más alto de abstracción describe sólo una parte de la base de datos completa.

Aunque el nivel lógico utiliza estructuras más simples, la complejidad se mantiene debido a la variedad de información almacenada en una base de datos grande.

Muchos usuarios del sistema de base de datos no necesitan toda esta información, sólo necesitan acceder a una parte de la base de datos. El nivel de vista de la abstracción existe para simplificar su interacción con el sistema.

1.6. Usuarios de una Base de Datos

Hay cuatro grupos de personas que intervienen en el entorno de un sistema de base de datos: el administrador de la base de datos, los diseñadores de la base de datos, los programadores de aplicaciones y los usuarios.

■ Diseñadores de la Base de Datos

- Encargado de grabar los propios Modelos de Datos
- Esquema de la Base de Datos
- Diseño lógico de la Base de Datos

■ Administrador de la Base de Datos

Encargado de:

- Monitorear el Performance
- Diseño físico de la base de datos y de su implementación.
- Herramientas Administrativas
 - Creacion de cuentas de usuarios
 - Objetos accedidos.
 - Matriz de autorizacion.
- Definir tiempos de respaldo
- Reorganización físico
- Llevar a cabo las tecnicas de recuperación

■ Programadores de Aplicaciones

Son los que se encargan de implementar los programas de aplicación que servirán a los usuarios finales. Estos programas son los que permiten consultar datos, insertarlos, actualizarlos y eliminarlos.

- Interfaces de los usuarios finales
 - Facilitar el acceso a ciertos "objetos" de la BD
- Interfaces para la gestión de la aplicaciones
 - Operaciones de escritura sobre altas ó bajas
 - IDE desarrollo (Java, .NET)
 - Lenguaje Scripts
 - Conectividad servidores de datos (API s)

1.6.1. Usuarios Finales

Los usuarios finales son los clientes de la base de datos, son las personas que requieren acceso a la base de datos para realizar consultas, actualizaciones e informes. Los usuarios se pueden clasificar en varias categorías:

- **Casuales** Estos acceden ocasionalmente a la base de datos, pero pueden necesitar una información diferente en cada momento.
 - Data Mining
 - Big Data
- **Principiantes - Paramétricos** Constituyen una parte considerable de los usuarios finales de los sistemas de bases de datos. Su labor principal gira entorno a la consulta y actualización constantes de la BD.
- **Sofisticados**
 - Experiencia en aplicaciones
 - Programadores de aplicación
 - DBA
 - Investigadores
- **Independientes (Stand Alone)**
 - Sistemas Escolares
 - Sistemas de prueba
 - Sin conectividad a otros nodos

1.7. DML vs DDL

■ DML

Data Manipulation Lenguaje, es un lenguaje que nos permite modificar los datos guardados.

Los tipos de acceso que tenemos disponibles son:

- Recuperar información
- Insertar nueva información
- Eliminar la información
- Modificación de la información

Una consulta o query es una sentencia que solicita la recuperación de información. La parte de un DML que implica recuperación de información se denomina lenguaje de consulta.

Aunque técnicamente incorrecto, solemos utilizar los términos lenguaje de consulta y lenguaje de manipulación de datos como sinónimos.

■ DDL

Data Definition Lenguaje, es Lenguaje de definición de datos.

Especificamos un esquema de base de datos mediante un conjunto de definiciones expresadas por un lenguaje especial denominado DDL.

El DDL también se utiliza para especificar propiedades adicionales de los datos.

Podemos simplificar en el sentido de SQL que:

DDL: Data Definition Language, mediante el cual puede definir nuevos objetos de base de datos, como Table, Views, Stored Procedures, etc. Algunos comandos comunes son:

- CREATE
- ALTER
- DROP
- etc...

DML: Data Manipulation Language, mediante el cual puede realizar cambios en los objetos creados anteriormente por DDLs. Algunos comandos comunes son:

- INSERT
- UPDATE
- DELETE
- SELECT
- etc...

1.8. Arquitectura del Sistema Gestor

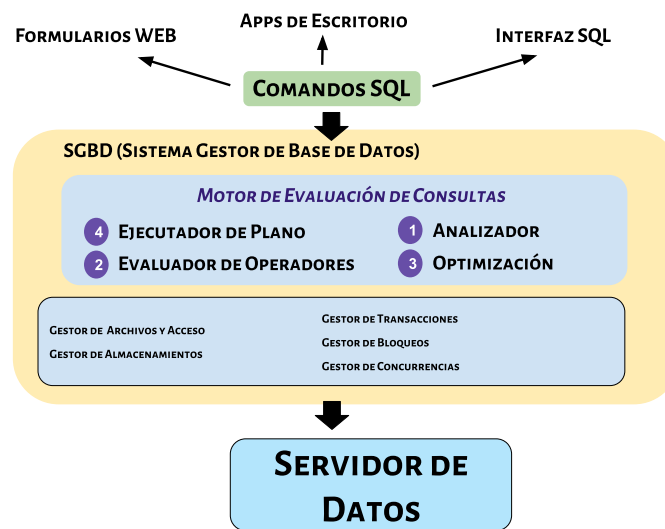
Podemos acceder al Sistema Gestor de nuestra Base de Datos de muchas maneras, desde formularios web, aplicaciones de escritorio e interpretes de SQL, todos se comunican con la Base de Datos mediante SQL.

■ Motor de Evaluación de Consultas

- Analizador: Este se encarga de analizar las sentencias SQL a nivel sintactico y lexico, así como una validación de que existan dichas relaciones y atributos. Es lo que mucha gente conoce como un compilador de DDL ó DML.
- Evaluador de Operaciones: Este se encarga de crear el árbol canónico, es decir, la forma formal que tendría el árbol necesario para acceder a la información.
- Optimizador: Se encarga de tomar el árbol canónico y optimizarlo para que se tenga que hacer la menor cantidad de operación a nivel lógico, este se conoce como árbol de consulta.
- Ejecutor de Planes: Se encarga de planear la mejor estrategia para ejecutar la consulta de tal manera que se optimize de manera física.

■ Gestores

- **Gestor de Transacciones:** Este es el que se encarga de organizar varias sentencias SQL para ejecutarlas como una transacción.
- **Gestor de Bloqueos:** Este es el que se encarga de ver si es que cierta relación esta bloqueada porque esta ocurriendo una transacción en ese momento.
Los bloqueos son una parte muy importante de este gesto.
Podemos definirlos en principalmente 2:
 - **Binario:** Es decir permite que una relación este o bien bloqueada o no para lectura o escritura.
 - **Múltiplos Niveles:** Es decir nos permite que existan bloqueos de lectura, escritura diferentes, ayudando a evitar las esperas si no son necesarias.
- **Gestor de Recuperación:** Este es el que se encarga de definir todas las técnicas de recuperación sobre la base, estas mismas las podemos definir en dos:
 - **Inmediatas:** Son las que podemos representar como rollback o rollforward
 - **Técnicas en Frío:** Usa una copia de seguridad
- **Gestor de Archivos y Métodos de Acceso** Todos los elementos, metadatos, logs, funciones, etc... tienen que estar almacenada de manera física.
Este se encarga de dependiendo de las queries como es que debe acceder de manera eficiente a los datos.
- **Gestor de Memoria Intermedia** Es como una RAM, nos permite guardar consultas bastante frecuentes para aumentar la velocidad sobre la RAM



1.9. Arquitecturas de las Bases de Datos

■ Cliente - Servidor

Mientras que el servidor es aquel que contiene toda la información importante, el cliente solo se encarga de solicitar servicios.

Podemos separar estos como:

- Cliente sin Disco: Aquel que puede acceder a la información pero no puede modificarlo.
- Cliente con Disco: Aquel que puede acceder a la información y puede modificarlo.

■ Multibase

Esta se caracteriza porque cada base de datos interna que tienes no se relaciona entre si, es decir, no hay comunación entre ellas.

Capítulo 2

Modelos de Datos

2.1. Sistema Entidad-Relación

Fue creado por Peter Pin-Shan Chen en 1976, llamado “The Entity Relationship Model Toward Unified View of Data”

Es un modelo conceptual, es decir, creado en lenguaje natural y se se ayuda de DER (gráficas) para hablar de las relaciones.

Permite construir el modelo conceptual de datos.

- Es una representación de la estructura y contenido de una base de datos.
- Es independiente del software (como el SMBD).
- Permite establecer las restricciones de la BD.
- Esta asociado al modelo de datos que es usado para implementar la BD.

Siendo una publicado por la ACM (Association for Computing Machinery), considera los siguientes puntos:

- Incorpora información semántica del mundo real.
- Introduce una técnica gráfica como herramienta para el diseño de bases de datos (Diagrama Entidad Relación).
- Es una representación lógica de los datos de una organización o un área de negocio.
- Normalmente es expresado como un DER, siendo este la representación gráfica del Modelo ER.

2.1.1. Entidades

Son objetos que existen en el mundo real y que son distinguibles (gracias a sus características) de otros objetos.

El tipo de entidad es el esquema que tiene la entidad en la base de datos, es decir son todos los atributos o características que definen a la entidad.

Piensa en las entidades como sustantivos. Ejemplos: cliente, estudiante, automóvil o producto.

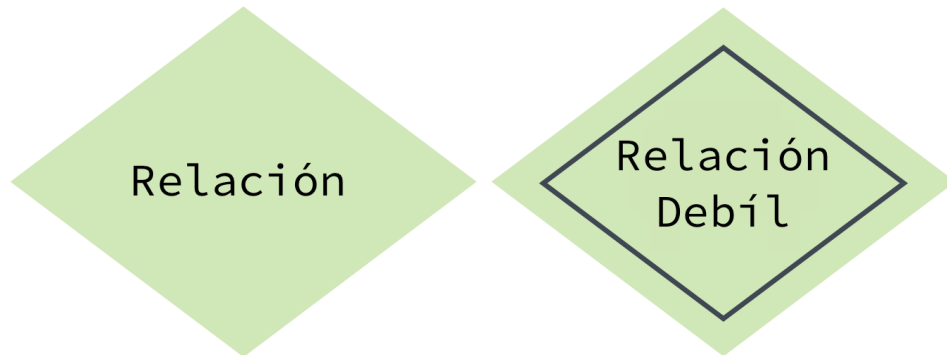


- Fuerte:
Es aquella entidad cuya existencia no depende de otras entidades.
- Debil:
Es aquella entidad cuya existencia depende de otras entidades.
 - Esta depende de de una o varias entidades fuertes
 - Una entidad debil jamas se debe asociar con una entidad debil
 - Una entidad debil no tiene identificador propio, pero si parciales
 - Usa una relación identificada para asociar una entidad fuerte con una debil
- Asociativa:
Se usa cuando tienes una relación y resulta que quieres usar esta relación como conexión (por ejemplo la relación entre alumno y clase) así que haces esa relación una entidad asociativa (por ejemplo para asociar esta nueva entidad con profesor).

2.1.2. Relaciones

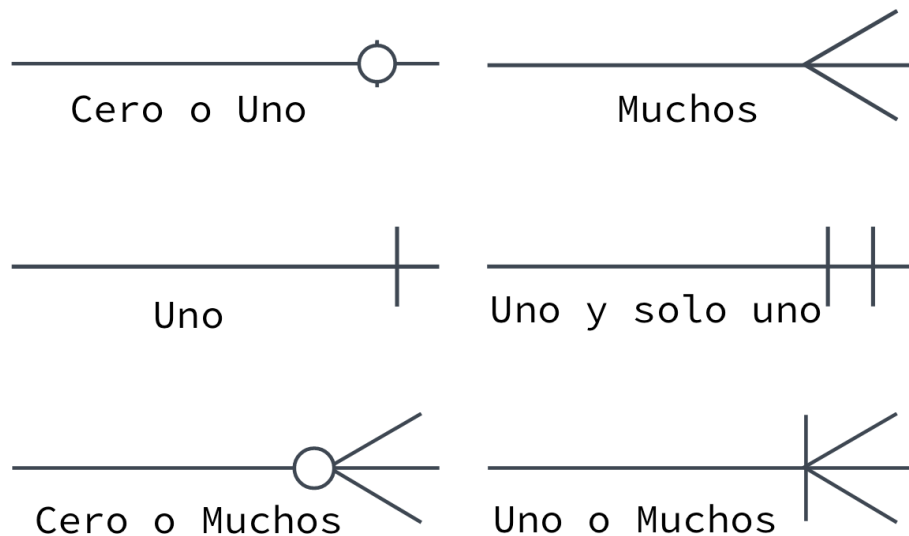
Esto es como las entidades actúan unas sobre otras o están asociadas entre sí. Piensa en las relaciones como verbos.

Por ejemplo, el estudiante nombrado puede registrarse para un curso. Las dos entidades serían el estudiante y el curso, y la relación representada es el acto de inscribirse, conectando las dos entidades de esa manera.



2.1.3. Cardinalidad

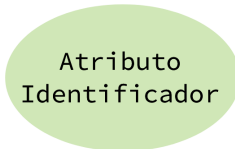
Nos dice la cantidad de atributos que se relacionan.



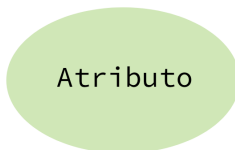
2.1.4. Atributos

Son las propiedades o características de una entidad

- **Identificador:** Es aquel atributo que nos permite identificar de forma unica cada una de las instancias de la entidad.



- **Simple:** Son aquellas propiedades atómicas que ya no se pueden subdividir



- **Derivado:** Es aquel que se puede calcular con la información que ya tenemos como la edad con la fecha de nacimiento



- **Multivalor:** Aquellos que tienen mas de un valor



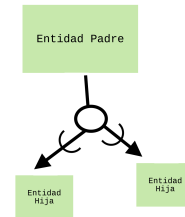
- **Compuestos :** Aquellos que estan formados por varios sub atributos

2.2. Sistema Entidad-Relación Extendido

El Modelo de Relación de Entidad Extendida es un modelo más complejo y de más alto nivel que extiende un diagrama E-R para incluir más tipos de abstracción y para expresar más claramente las restricciones.

Todos los conceptos de un diagrama E-R están incluidos en el modelo EE-R.

Vamos a ver las diferencias más importantes con respecto a nuestro clásico amigo el diagrama entidad relación.



■ Herencia: Super y Subclases

Como el nombre sugiere, podemos tener entidades que consideramos padres y que heredan a otras entidades hijas.

Por ejemplo podemos tener una entidad empleado puede tener subentidades que se dividen por el trabajo que hacen, por ejemplo diseñadores, gerentes y ayudantes.

Todas nuestras subentidades van a heredar las propiedades o atributos de nuestra superentidad.

Esto nos permite que nuestra entidad padre sea una idea general y que nuestras subclases sean especializaciones de la misma, por esto decimos que esta clase de diagramas permite la especialización.

Es también importante remarcar que alguna de nuestras entidades puede heredar de una o mas entidades

■ Generalización y Especialización

Como habíamos hablando antes, esto nos permite que si tenemos una gran cantidad de entidades con atributos comunes podemos crear una entidad general y hacer que las demás hereden de ella, con esto también propiciamos que nuestras entidades que heredan se puedan especializar.

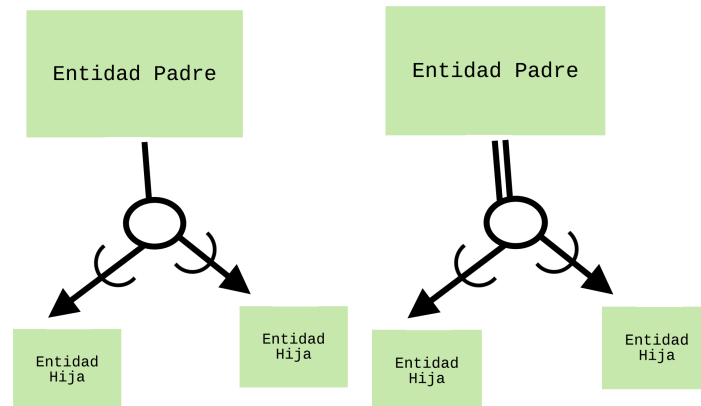
■ Restricciones También es común que se les conozca como superclass constraint

- Entidades Disconjuntas
- Entidades Superpuestas

■ Discriminadores de Superentidad

Un discriminador de entidades es un atributo que indica el subtipo de una entidad.

Es decir podemos indicar con este diagrama si las divisiones que hacemos son parciales (es decir, es posible que una instancia de la superclase pueda existir sin tener que ser parte de una subentidad) o total en la que decimos que cualquier instancia tiene que pertenecer a alguna subentidad



2.3. Sistema Relacional

Las bases de datos relacionales facilitan evitar la duplicidad de información e inconsistencias de datos. La mayoría de los sistemas de bases de datos utilizados hoy en día son relacionales.

En el modelo relacional, los datos se dividen en diferentes **tables**. Una tabla funciona como una matriz o una hoja de cálculo.

Normalmente, las columnas imponen un tipo de datos que pueden contener. Las columnas también pueden especificar otras restricciones: si es obligatorio que las filas tengan un valor en esa columna, si el valor de la columna debe ser único en todas las filas y cosas así.

Las columnas se denominan más comúnmente atributos. Si una columna sólo permite números enteros, decimos que es un campo entero. Diferentes tablas utilizan diferentes tipos de campos.

La organización de una tabla de base de datos viene dada por sus campos y las restricciones que imponen.

Todas las entradas de datos son filas y el sistema de base de datos no aceptará una fila en una tabla si viola el esquema de la tabla. Esa es una gran limitación del modelo relacional. Cuando las características de los datos varían demasiado, la adaptación de los datos a un esquema fijo puede ser problemática. Pero si está trabajando con datos de estructura homogénea, un esquema fijo le ayudará a garantizar que los datos son válidos.

2.3.1. Relaciones

Para evitar la redundancia de datos la clave esta en crear diversas tablas e ir relacionandolas.

Para lograr esto tenemos que crear relaciones, y es imposible hablar de relaciones sin hablar de llaves:

- **Primary Key:** La idea de una llave primaria es crear un atributo que tiene, tiene pero tiene que ser único y es recomendable que casi no cambie.
- **Foreign Key:** La idea de una llave secundaria es crear un atributo que tiene que alguna otra tabla sea la llave primaria. Eso es todo.

Capítulo 3

SQL

3.1. SQL como Lenguaje

SQL **no** es tan potente como una máquina universal de Turing.

Es decir, hay algunos cálculos que son posibles utilizando un lenguaje de programación de propósito general, pero no son posibles con SQL. SQL también no admite acciones como la entrada de usuarios, la salida a las pantallas ó la comunicación a través de la red. Dichas computaciones y acciones deben escribirse en un lenguaje principal (C, C++ ó Java) con consultas SQL incorporadas que acceden a los datos de la base de datos

Los programas de aplicación son programas que se utilizan para interactuar con la base de datos de esta manera.

Parte II

Parte Practica

Capítulo 4

Queries en SQL

4.1. SELECT FROM

Empecemos por la sentencia más básica para un Query en SQL:

```
1 SELECT FielName FROM TableName;
```

Ahora empecemos por aquí:

4.1.1. Distint vs All

- Distint (Default): Esta solo muestra solo los atributos diferentes

```
1 SELECT DISTINCT FielName FROM TableName;
```

- All: Esta muestra todos los atributos, incluso duplicados

```
1 SELECT ALL FielName FROM TableName;
```

4.2. Funciones de Agregación

Las funciones de agregación son funciones que toman una colección como entrada y regresan un simple valor. SQL ofrece estas 5 funciones por default:

- “AVG()”: Promedio
- “MIN()”: Minimo
- “MAX()”: Máximo
- “COUNT()”: Cuenta la cantidad
- “SUM()”: Suma todo el resultado

4.3. WHERE

Esta nos permite seleccionar solo las filas en el resultado del query que satisface nuestros criterios.

```
1 SELECT FielName FROM TableName
2 WHERE
3     (TableName.Thing1 < Something) AND
4     (TableName.Thing2 = "Some_Strings");
```

Tenemos a nuestra disposición los siguientes comparadores:

- Lógicos: AND, OR, NOT
- Aritmeticos: <, >, <=, >=, =, <>

4.4. Comparación de Strings

Podemos comparar strings gracias a lo que conocemos como comodines usando el comando “LIKE”

- “%”: Que sirve como comodín para cualquier cantidad de caracteres
- “_”: Que sirve como comodín para un caracter

4.5. Orden de las Tuplas

Podemos controlar el orden en que se nos retornan las tuplas del query que estamos haciendo muy facilmente con la instructiva ORDER BY.

Por default se usa la opción “ASC” que las ordena de manera ascendente mientras que “DESC” la ordena de manera descendente.

Por default esta se coloca la última línea de manera implícita:

```
1 SELECT DISTINCT FielName1, FielName2, FielName3 FROM TableName
2 ORDER BY FielName1 ASC, FielName2 ASC, FielName3 ASC;
```

También podemos simplemente poner el número en que las pusimos para indicar el orden:

```
1 SELECT DISTINCT FielName1, FielName2, FielName3 FROM TableName
2 ORDER BY 1,2, 3;
```