



Modelo Relacional

Bases de datos I – UTN FRRO

Giuliano Crenna

Historia y Origen del Modelo Relacional: Edgar F. Codd y la Revolución en Bases de Datos

El modelo relacional es un concepto fundamental en el campo de las bases de datos, y su creación marcó un antes y un después en la manera en que se estructuran y manipulan los datos. Este modelo fue propuesto por el matemático y científico de la computación británico Edgar Frank Codd en 1970. Su trabajo no solo revolucionó el almacenamiento y acceso a datos, sino que también sentó las bases para la evolución de sistemas de bases de datos modernas, influyendo en la creación de SQL y en el desarrollo de las bases de datos relacionales que son ampliamente utilizadas hoy en día.

Antecedentes: La Necesidad de una Nueva Estructura

Antes del surgimiento del modelo relacional, las bases de datos se organizaban utilizando modelos jerárquicos y de red. Estos modelos eran complicados, difíciles de manejar y no ofrecían la flexibilidad necesaria para adaptarse a los crecientes volúmenes y complejidades de los datos. En los años 60, con el auge de las computadoras y la digitalización de la información, la gestión eficiente de grandes cantidades de datos se convirtió en un desafío crítico.

Modelos Precedentes:

- **Modelo Jerárquico:** Organiza los datos en una estructura tipo árbol, donde cada nodo tiene un solo padre y puede tener múltiples hijos. Aunque útil en ciertos casos, este modelo presentaba problemas de redundancia y dificultad para representar relaciones más complejas.
- **Modelo de Red:** Permite relaciones más complejas que el modelo jerárquico, con nodos que pueden tener múltiples padres. Sin embargo, su complejidad también dificultaba la manipulación de datos, requiriendo que los programadores trabajaran directamente con punteros y referencias.

Principios Clave del Modelo Relacional:

- **Tablas (Relaciones):** En lugar de estructuras jerárquicas o en red, los datos se organizarían en tablas bidimensionales, donde cada tabla (o relación) representa una entidad del mundo real.
- **Tuplas y Atributos:** Las filas en las tablas se denominan tuplas y representan instancias de entidades, mientras que las columnas (atributos) describen las propiedades de esas entidades.
- **Llaves Primarias y Foráneas:** Introdujo el concepto de llaves primarias para identificar de manera única cada fila en una tabla, y llaves foráneas para establecer relaciones entre tablas.
- **Independencia de Datos:** Los usuarios pueden interactuar con los datos a un nivel abstracto sin necesidad de conocer los detalles de su

almacenamiento físico, permitiendo la independencia lógica y física de los datos.

- Álgebra Relacional: Propuso un conjunto de operaciones algebraicas (como selección, proyección y unión) para manipular las tablas y extraer información.

Impacto y Evolución del Modelo Relacional

El modelo relacional tuvo un impacto profundo en el desarrollo de la informática y las bases de datos. Aunque al principio enfrentó resistencia, especialmente dentro de IBM, debido a la complejidad de su implementación, eventualmente se reconoció como un avance crucial. A mediados de los años 70, IBM comenzó a trabajar en un sistema de gestión de bases de datos basado en el modelo relacional, conocido como System R, que a su vez influyó en el desarrollo del lenguaje SQL (Structured Query Language).

El trabajo de Codd llevó a la creación de sistemas de bases de datos relacionales que hoy son la base de la mayoría de las aplicaciones de software en todo el mundo, desde sistemas financieros hasta redes sociales. El impacto de su propuesta es tal que en 1981, Codd fue galardonado con el prestigioso Premio Turing, uno de los mayores reconocimientos en el campo de la computación, por sus contribuciones al modelo relacional.

Definición del Modelo Relacional

El modelo relacional es una de las bases más importantes en el campo de la ciencia de datos y bases de datos. Introducido por Edgar F. Codd en 1970, este modelo representa una forma estructurada y matemática de organizar, almacenar y manipular datos en sistemas de bases de datos. Su simplicidad, flexibilidad y capacidad para manejar grandes volúmenes de información lo han convertido en el estándar dominante en la industria.

Definición del Modelo Relacional

El modelo relacional se basa en la idea de que todos los datos en una base de datos deben representarse mediante relaciones, que son esencialmente tablas bidimensionales. Cada tabla está compuesta por filas (conocidas como tuplas) y columnas (conocidas como atributos), y cada celda de la tabla contiene un valor de dato atómico.

Elementos Clave del Modelo Relacional

1. Relaciones (Tablas):

- En el modelo relacional, una relación se representa como una tabla, donde cada fila de la tabla corresponde a una entidad del mundo real y cada columna corresponde a un atributo de esa entidad.
- Por ejemplo, una tabla llamada "Clientes" podría tener columnas como "ID", "Nombre", "Correo Electrónico" y "Fecha de Registro". Cada fila de esta tabla representaría un cliente individual.

2. Tuplas (Filas):

- Una tupla es una fila en una tabla que representa una instancia específica de una entidad. Cada tupla contiene un conjunto de valores, uno por cada atributo.
- Por ejemplo, una tupla en la tabla "Clientes" podría contener los valores "001", "Juan Pérez", "juan.perez@example.com", "2024-08-14", representando a un cliente específico.

3. Atributos (Columnas):

- Un atributo es una columna en una tabla que describe una propiedad o característica de las entidades representadas en la tabla.
- Los atributos son definidos por un nombre y un dominio, que es el conjunto de valores permitidos para ese atributo.

4. Dominios:

- Un dominio es el conjunto de valores posibles que un atributo puede tener. Por ejemplo, el dominio de un atributo "Edad" podría ser el conjunto de todos los números enteros positivos.

5. Llaves:

- Llave Primaria: Es un atributo o conjunto de atributos que identifica de manera única cada tupla en una tabla. Una tabla no puede tener dos tuplas con el mismo valor de llave primaria.
- Llave Foránea: Es un atributo en una tabla que crea una relación con la llave primaria de otra tabla, permitiendo la vinculación de datos entre diferentes tablas.

6. Integridad Relacional:

- El modelo relacional asegura la integridad de los datos mediante reglas como la integridad referencial, que asegura que las llaves foráneas correspondan a valores existentes en la tabla relacionada.



Operaciones en el Modelo Relacional

El modelo relacional utiliza un conjunto de operaciones para manipular y consultar los datos en las tablas. Estas operaciones forman parte del álgebra relacional, que es un lenguaje formal basado en la teoría de conjuntos.

1. **Selección:** Extrae las tuplas que cumplen con una condición específica.
2. **Proyección:** Selecciona un subconjunto de columnas de una tabla, eliminando las restantes.
3. **Unión:** Combina los resultados de dos tablas con la misma estructura.
4. **Diferencia:** Encuentra las tuplas que están en una tabla pero no en otra.
5. **Producto Cartesiano:** Combina todas las tuplas de una tabla con todas las tuplas de otra.
6. **Unión Natural:** Combina dos tablas basándose en la igualdad de valores en los atributos compartidos.

Ventajas del Modelo Relacional

- **Simplicidad:** La estructura tabular es intuitiva y fácil de entender.
- **Flexibilidad:** Permite la adición y eliminación de datos sin afectar la integridad del sistema.
- **Independencia de Datos:** El modelo relacional abstrae la manera en que se almacenan los datos, lo que facilita su manipulación sin conocer detalles físicos del almacenamiento.
- **Consistencia e Integridad:** Las reglas de integridad y las operaciones relacionales aseguran que los datos sean consistentes y correctos.

Conceptos Fundamentales del Modelo Relacional

El modelo relacional, propuesto por Edgar F. Codd, se basa en conceptos fundamentales que permiten organizar y manipular datos de manera estructurada y eficiente. Entre estos conceptos se encuentran las relaciones, atributos, tuplas y dominios, que juntos forman la base del modelo relacional y permiten la creación de sistemas de bases de datos robustos y flexibles.

Relaciones: Definición de Tablas

En el modelo relacional, una relación es la representación de un conjunto de entidades en forma tabular, es decir, como una tabla bidimensional. Cada tabla en una base de datos relacional se denomina "relación" y se utiliza para almacenar datos relacionados de manera estructurada.

- **Estructura de la Relación:**

- Una relación se compone de filas y columnas. Las filas representan instancias específicas de una entidad, mientras que las columnas representan los atributos o propiedades de esas instancias.
- Por ejemplo, una tabla llamada "Estudiantes" puede tener columnas como "ID", "Nombre", "Edad" y "Carrera", donde cada fila representa a un estudiante específico.

- **Relaciones en la Práctica:**

- Las relaciones permiten almacenar datos de manera que las conexiones entre diferentes conjuntos de datos se puedan gestionar fácilmente. Por ejemplo, en un sistema de gestión universitaria, una relación "Estudiantes" puede estar relacionada con una relación "Cursos" mediante una llave foránea, lo que permite determinar qué cursos está tomando cada estudiante.

- **Propiedades de las Relaciones:**

- Una relación no tiene un orden intrínseco en sus filas o columnas; el ordenamiento se puede realizar de acuerdo con las necesidades de la consulta.
- Cada relación debe tener un conjunto único de columnas (atributos), y cada fila debe ser única, lo que significa que no puede haber duplicados.

Atributos: Columnas en las Tablas

Los atributos son las columnas de una relación (tabla) y describen las características o propiedades de las entidades que se almacenan en esa tabla. Cada atributo tiene un nombre y está asociado con un dominio, que define los valores posibles que puede tomar.

- **Definición de Atributos:**

- Un atributo describe una propiedad específica de la entidad que la tabla representa. Por ejemplo, en una tabla "Empleados", los atributos pueden incluir "Nombre", "Apellido", "Número de Seguro Social", "Posición", y "Salario".
- Los atributos son los que definen la estructura de los datos que se almacenarán en cada fila de la tabla.

- **Dominio del Atributo:**

- Cada atributo tiene un dominio asociado, que es el conjunto de valores permitidos para ese atributo. Por ejemplo, el atributo "Edad" puede tener un dominio de números enteros positivos, mientras que

el atributo "Correo Electrónico" puede tener un dominio de cadenas de texto con un formato específico.

- El dominio asegura que los datos sean válidos y consistentes con el propósito del atributo.

- **Tipos de Atributos:**

- Los atributos pueden ser de diferentes tipos, como numéricos, cadenas de texto, fechas, booleanos, entre otros. La elección del tipo de atributo depende del tipo de información que se pretende almacenar.

Tuplas: Filas en las Tablas

Las tuplas representan las filas de una tabla en el modelo relacional. Cada tupla es una instancia de la entidad representada por la tabla y contiene un conjunto de valores, uno por cada atributo.

- **Definición de Tuplas:**

- Una tupla es una secuencia ordenada de valores, donde cada valor corresponde a un atributo de la relación. Por ejemplo, en una tabla "Productos", una tupla puede representar un producto específico, con valores como "001", "Laptop", "1500", "Electrónica", donde cada valor se asocia con atributos como "ID", "Nombre", "Precio" y "Categoría".
- Cada tupla debe ser única dentro de una relación, lo que significa que no puede haber dos filas idénticas en la misma tabla.

- **Identificación de Tuplas:**

- En una tabla relacional, las tuplas suelen identificarse de manera única mediante una **llave primaria**. Una llave primaria es un atributo o un conjunto de atributos cuyos valores son únicos en cada tupla. Por ejemplo, en una tabla "Clientes", el atributo "ID de Cliente" podría servir como llave primaria.
- Las tuplas permiten que la información sobre diferentes instancias de una entidad se almacene de manera organizada y estructurada.

- **Orden de las Tuplas:**

- El modelo relacional no impone un orden específico para las tuplas dentro de una relación. El ordenamiento puede realizarse en función de las necesidades del usuario o del sistema, como al ejecutar consultas específicas.

Dominios: Conjunto de Valores Permitidos para un Atributo

El dominio en el modelo relacional se refiere al conjunto de valores posibles que un atributo puede tomar. Los dominios son cruciales para mantener la integridad y la consistencia de los datos dentro de la base de datos.

- **Definición de Dominio:**
 - Un dominio es un conjunto bien definido de valores posibles para un atributo. Por ejemplo, el dominio para el atributo "Estado Civil" en una tabla "Personas" podría ser el conjunto {"Soltero", "Casado", "Divorciado", "Viudo"}.
 - El dominio define el tipo de datos que un atributo puede almacenar, como números enteros, decimales, cadenas de texto, fechas, etc.
- **Importancia del Dominio:**
 - Los dominios son esenciales para la validación de datos, asegurando que los valores ingresados en los atributos sean válidos y apropiados para el contexto. Por ejemplo, un dominio para un atributo "Edad" podría restringir los valores a números enteros positivos.
 - También ayudan a evitar errores y garantizar la consistencia en el almacenamiento y la manipulación de datos.
- **Restricciones de Dominio:**
 - Además del tipo de datos, un dominio puede incluir otras restricciones, como rangos específicos (e.g., edades entre 0 y 120) o patrones específicos para cadenas de texto (e.g., direcciones de correo electrónico con un formato particular).

Operaciones Fundamentales del Modelo Relacional

El modelo relacional de bases de datos se basa en un conjunto de operaciones fundamentales que permiten manipular y consultar datos de manera eficiente y estructurada. Estas operaciones, que incluyen la selección, proyección, unión, diferencia, producto cartesiano y unión natural, son esenciales para trabajar con datos organizados en tablas (relaciones). A continuación, exploramos cada una de estas operaciones de manera detallada, pero sin recurrir a fórmulas matemáticas.

1. Selección

La selección es una operación que permite filtrar las filas (tuplas) de una tabla basándose en una condición específica. Esta operación devuelve una nueva tabla que solo contiene las filas que cumplen con la condición dada.



- **Propósito:** Extraer datos relevantes de una tabla eliminando las filas que no cumplen con los criterios establecidos.
- **Cómo Funciona:** Se especifica una condición, y la selección devuelve todas las filas que satisfacen esa condición. Por ejemplo, si tienes una tabla de empleados, puedes usar la selección para obtener solo los empleados que trabajan en un departamento específico o que tienen un salario superior a cierto monto.
- **Ejemplo:** De una tabla "Empleados", seleccionar solo aquellos con una edad mayor a 30 años.

2. Proyección

La proyección es una operación que se utiliza para seleccionar un subconjunto de columnas (atributos) de una tabla, eliminando las columnas que no son necesarias para la consulta o análisis.

- **Propósito:** Simplificar los datos mostrando solo la información relevante y eliminando detalles innecesarios.
- **Cómo Funciona:** Se eligen las columnas que se desean visualizar, y la proyección crea una nueva tabla que incluye solo esas columnas, manteniendo todas las filas originales.
- **Ejemplo:** De una tabla "Estudiantes" con columnas como "ID", "Nombre", "Edad" y "Carrera", se podría proyectar solo "Nombre" y "Carrera" para obtener una lista de estudiantes con su respectiva carrera.

3. Unión

La unión es una operación que combina dos tablas con el mismo esquema, es decir, con las mismas columnas, en una sola tabla. Esta operación incluye todas las filas de ambas tablas, eliminando las duplicadas.

- **Propósito:** Combinar datos de diferentes fuentes que tienen la misma estructura para obtener una visión más completa.
- **Cómo Funciona:** La unión toma todas las filas de las dos tablas y las incluye en una nueva tabla. Si hay filas duplicadas (es decir, exactamente iguales en ambas tablas), solo se mantendrá una.
- **Ejemplo:** Si tienes una tabla de "Clientes" de una tienda física y otra de clientes de una tienda en línea, la unión combinaría ambos conjuntos de datos en una sola tabla, mostrando todos los clientes sin repetir ninguno.

4. Diferencia

La diferencia es una operación que compara dos tablas con el mismo esquema y devuelve las filas que están en la primera tabla pero no en la segunda.



- **Propósito:** Identificar datos únicos en una tabla en comparación con otra.
- **Cómo Funciona:** La operación de diferencia resta las filas de la segunda tabla de la primera. La nueva tabla resultante contendrá solo aquellas filas que están presentes en la primera tabla y no en la segunda.
- **Ejemplo:** Si tienes una tabla de "Productos Vendidos" y otra de "Productos en Inventario", la diferencia entre "Inventario" y "Vendidos" te dará una lista de productos que no se han vendido.

5. Producto Cartesiano

El producto cartesiano es una operación que toma dos tablas y crea una nueva tabla combinando cada fila de la primera tabla con cada fila de la segunda tabla.

- **Propósito:** Generar todas las combinaciones posibles de datos entre dos conjuntos de datos diferentes.
- **Cómo Funciona:** Cada fila de la primera tabla se empareja con cada fila de la segunda tabla, resultando en una tabla que contiene todas las combinaciones posibles. Este método es útil para preparar datos para análisis más complejos, aunque puede generar un gran número de filas.
- **Ejemplo:** Si tienes una tabla de "Clientes" y otra de "Productos", el producto cartesiano crearía una tabla con todas las combinaciones posibles de clientes y productos, mostrando qué productos podría comprar cada cliente.

6. Unión Natural

La unión natural es una operación que combina dos tablas en función de las columnas que tienen en común, fusionando aquellas filas donde los valores coinciden en dichas columnas.

- **Propósito:** Relacionar y combinar datos de tablas diferentes que comparten un atributo común, eliminando la redundancia.
- **Cómo Funciona:** La operación busca las columnas comunes entre las dos tablas y une las filas donde los valores de estas columnas coinciden, creando una nueva tabla. Las columnas duplicadas que resultan de la combinación se eliminan.
- **Ejemplo:** Si tienes una tabla "Pedidos" con columnas "ID_Cliente" y "ID_Pedido", y otra tabla "Clientes" con "ID_Cliente" y "Nombre", la unión natural combinaría ambas tablas para mostrar qué pedidos corresponden a qué clientes, incluyendo solo una columna "ID_Cliente".



Bibliografía:

- **Codd, E. F. (1970).** "A Relational Model of Data for Large Shared Data Banks". *Communications of the ACM*, 13(6), 377-387.
- **Elmasri, R., & Navathe, S. B. (2016).** *Fundamentals of Database Systems* (7th ed.). Pearson.
- **Silberschatz, A., Korth, H. F., & Sudarshan, S. (2019).** *Database System Concepts* (7th ed.). McGraw-Hill Education.