

BASES DE DATOS II

Tema 2

Bases de Datos Relacionales

Víctor Valenzuela Ruz
INACAP Copiapó

Modelo Relacional (1/2)

- RDBM's nacieron en 1970 por el trabajo de E.F. **Codd**.
- Están basadas en relaciones (**tablas**) como estructuras de almacenamiento.
- Codd propone un modelo de datos basados en la teoría de las **relaciones**, en donde los datos se estructuran lógicamente en forma de relaciones (tablas), siendo un objetivo fundamental del modelo: mantener la independencia de esta estructura lógica respecto al modo de almacenamiento y a otras características de tipo físico.

Modelo Relacional (2/2)

- DBMS han sido diseñadas para aplicaciones de negocios (**transaccionales**).
- Estandarizaron el lenguaje de manipulación, usando **SQL** (*Structured Query Language*).
- Han madurado sobre los últimos 15 años, y aún son populares.
- Alta eficiencia en el procesamiento transaccional (**OLTP**).
- Permiten trabajo **concurrente** (muchos usuarios en línea)
- Mantienen restricciones de **integridad** (simples), preservando la **consistencia** de los datos.

Objetivos del Modelo Relacional

- **Independencia Física.** Es decir, que el modo en que se almacenan los datos no influya en su manipulación lógica y, por tanto, los usuarios que acceden a esos datos no tengan que modificar sus programas por cambios en el almacenamiento físico.
- **Independencia lógica.** Esto es, que el añadir, eliminar o modificar objetos de la base de datos no repercuta en los programas y/o usuarios que están accediendo a subconjuntos parciales de los mismos.
- **Flexibilidad.** En el sentido de poder presentar a cada usuario los datos de la forma en que éste prefiera.
- **Uniformidad.** Las estructuras lógicas de los datos presentan un aspecto uniforme, lo que facilita la concepción y manipulación de la base de datos por parte de los usuarios.
- **Sencillez.** Las características anteriores, así como unos lenguajes de usuario muy sencillos, producen como resultado que el modelo de datos relacional sea fácil de comprender y de utilizar por parte del usuario final.

Definiciones Relacionales

- Existe una serie de términos utilizados en el modelo relacional que se requiere definir:
 - **Relación** : Conjunto de filas. Pude asociarse a lo que se conoce como tabla, con ciertas propiedades.
 - **Tupla** : Corresponde a una fila de esa tabla. Al número de tuplas se denomina **cardinalidad**.
 - **Atributo**: Se refiere a una columna de esa tabla. La cantidad de atributos determina el **grado** de la relación.
 - **Clave Primaria**: Identificador único para la tabla. Se compone de una columna o de una combinación de ellas. Nunca existen dos filas de la misma tabla con el mismo valor de clave primaria.
 - **Dominio**: Es una colección de valores, de los cuales uno o más atributos obtienen sus valores reales.

Dominio

- Se define dominio como “un conjunto de valores escalares de donde extraen sus valores los atributos de una relación”. Los valores escalares representan la menor unidad semántica de información.
- ¿Cuál es la importancia de los dominios?
Los dominios tienen una importancia semántica, puesto que restringen las comparaciones. Los dominios no sólo definen los valores permitidos, sino además las comparaciones “permitidas”.

Relación (1/2)

- La cabecera está compuesta por un conjunto fijo de **atributos**, o en términos más precisos de pares atributos-dominio:

$$\{ (A_1:D_1), (A_2:D_2), \dots, (A_n:D_n) \}$$

tales que cada atributo A_j corresponde a uno y solo uno de los dominios subyacentes D_j .

- Una relación R , sobre un conjunto de dominios D_1, D_2, \dots, D_n (no necesariamente todos distintos), se compone de dos partes, una cabecera y un cuerpo.

Relación (2/2)

- El cuerpo está formado por un conjunto de **tuplas**, el cual varía con el tiempo. Cada tupla está compuesta por un conjunto de pares atributo-valor :

$$\{ (A_1:v_1), (A_2:v_2), \dots, (A_n:v_n) \}$$

para cada una de estas tuplas hay uno de estos pares atributo valor.

Propiedades de las Relaciones

- Las propiedades de las relaciones son un consecuencia de su definición:
 - No existen tuplas repetidas . Por tratarse, el cuerpo de una relación, de un conjunto matemático estos no admiten elementos repetidos. Como corolario aparece la clave primaria.
 - Las tuplas no están ordenadas (de arriba hacia abajo).
 - Los atributos no están ordenados (de izquierda a derecha)
 - Todos los valores de los atributos son atómicos. Es equivalente a decir las relaciones no contienen grupos repetitivos y en este caso estarían normalizadas.

Tablas Base y Derivadas

- **Tablas Base**

Son tablas creadas vía un comando del lenguaje de definición de datos, los datos asociados a estas tablas son almacenados permanentemente en memoria secundaria.
- **Tablas Derivadas**

Constituyen derivaciones de las tablas básicas, obtenidas mediante la utilización de comandos de un lenguaje de manipulación de datos. Los datos de este tipo de tabla tienen una vida temporal y normalmente constituyen datos redundantes.

Vistas o *Views*

- Las vistas también se denominan tablas virtuales. Las tuplas que compondrán las vistas serán generadas cada vez que la vista requiera ser utilizada.
- Los datos que componen una vista provienen de la ejecución de algunos comandos del lenguaje de manipulación de datos.
- Para crear una vista se utiliza un algoritmo especialmente definido para este fin, el cual es ejecutado cada vez que se utiliza la vista y es almacenado en memoria secundaria, este procedimiento es absolutamente transparente ara el usuario.
- Existen dos razones principales para utilizar Vistas:
 - Calcular valores en función de datos almacenados en las tablas de la base.
 - Restringir acceso tanto a filas como columnas de una tablas para determinados usuarios.

Reglas de Integridad

- El modelo relacional define dos reglas generales para mantener la integridad de los datos :
 - **Integridad de la entidad:** Una clave primaria no puede aceptar valores nulos.
 - **Integridad de las referencias:** Todo valor definido como clave foránea debe tener su correspondencia con un valor de clave primaria en la relación referenciada.

Base de Datos Relacional (1/2)

Definición

Conociendo los conceptos antes presentados, podemos definir una base de datos relacional (BDR) como “un conjunto de relaciones normalizadas* de distinto grado”.

* : Normalizada en primera forma normal.

Base de Datos Relacional (2/2)

- El modelo relacional es uno de los modelos de bases de datos más utilizados y que presenta una fuerte base teórica.
- Existe una gran variedad de Sistemas de Administradores de Bases de Datos en el mercado, basados en este modelo: Oracle, SyBase, Informix, Ingres, DB2, Microsoft SQL Server, entre otros.
- El modelo relacional suele dividirse en tres partes :
 - Estructura de datos (Relación)
 - Integridad
 - Manipulación

Componentes de un SABDR

- Un Sistema Administrador de Base de Datos (SABD) posee diversos componentes entre los que se cuentan:
 - Catálogo del sistema
 - Un lenguaje de datos (definición y manipulación)

Lenguaje de Consulta Relacional

- El modelo relacional considera un lenguaje de datos que permite la definición y manipulación de datos.
- El lenguaje de manipulación de datos (DML) permite la obtención de información, mediante operadores de consulta que se basan en el álgebra relacional.
- A su vez, el lenguaje de definición de datos (DDL) permite la definición de esquemas relacionales (creación de tablas, vistas, privilegios, etc.)

Álgebra Relacional

- El álgebra relacional consiste en un conjunto de **operadores** de alto nivel que operan sobre relaciones. Cada uno de estos operadores toma una o dos relaciones como entrada y produce una nueva relación de salida.
- Codd define un conjunto de ocho operadores clasificados en dos grupos:
 - Las **operaciones tradicionales** de conjuntos unión, intersección, diferencia y producto cartesiano.
 - Los **operadores especiales** de restricción, proyección, reunión y división.

Operaciones Tradicionales

- **Unión**: Construye una relación formada por todas las tuplas que aparecen en cualquiera de las dos relaciones especificadas
- **Intersección**: Construye una relación formada por aquellas tuplas que aparezcan en las dos relaciones especificadas
- **Diferencia**: Construye una relación formada por todas aquellas tuplas de la primera relación que no aparezcan en la segunda de las dos relaciones especificadas
- **Producto cartesiano**: A partir de dos relaciones especificadas, construye una relación que contiene todas las combinaciones posibles de tuplas, una de cada una de las relaciones.

Operadores Especiales

- **Restricción:** Extrae las tuplas especificadas de una relación dada (restringe la relación solo a las tuplas que satisfagan una condición especificada)
- **Proyección:** Extrae los atributos especificados de una relación dada
- **Unión:** A partir de dos relaciones especificadas, construye una relación que contiene todas las posibles combinaciones de tuplas, una de cada una de las dos relaciones, tales que las dos tuplas participantes en una combinación dada satisfagan alguna condición especificada
- **División:** Toma dos relaciones, una binaria y una unaria y construye una relación formada por todos los valores de un atributo de la relación binaria que concuerdan (en el otro atributo) con todos los valores de la relación unaria.

Lenguaje de Consulta Estructurado (SQL)

- SQL (*Structured Query Language*) es un lenguaje que permite la definición y la manipulación de una base de datos relacional.
- Se ha establecido como el lenguaje estándar de bases de datos relacionales, utilizándose en la mayoría de los SABDR como Oracle, Informix, Sybase, DB2, SQLServer, etc.
- Aunque las versiones de SQL utilizadas en productos difiere en aspectos del lenguaje, podría hablarse de la existencia de un estándar.

Características del SQL

- SQL se puede usar en forma interactiva o inmerso en lenguajes de programación de propósito general como Cobol, C++ , Power Builder, Centura, Visual Basic, etc.
- SQL se compone de:
 - DDL (lenguaje de definición de datos) : Define y modifica esquemas relacionales, crear tablas, crear vistas, definir accesos, etc.
 - DML (lenguaje de manipulación de datos) : Basado en el álgebra relacional permite la recuperación de información.

Conclusiones

- El modelo relacional se presenta como uno de los modelos más utilizados para la construcción de bases de datos.
- Su poderío se centra en que utiliza una sola estructura de datos, la relación, lo que le otorga gran flexibilidad, uniformidad y sencillez.
- SQL es el lenguaje estándar de las SABDR y permite definir y manipular tablas , siendo utilizado tanto en forma interactiva como inmerso en aplicaciones.

Auto-evaluación de Unidad

- Justifique la integridad y consistencia en el modelo relacional.
- ¿Cuál es la importancia de los dominios en un SABDR?
- Dé un ejemplo de Integridad en la Entidad y uno de Integridad de la Referencia.