

Normalización Adicional



Universidad Autónoma
de Entre Ríos

Introducción

- En la presentación anterior vimos como resolver la mayoría de los problemas que pueden presentar normalmente las relaciones.
- En esta presentación veremos casos particulares de relaciones que siguen sufriendo problemas. Estos casos son bastante raros de encontrarlos en la práctica.
- Son conocidos como casos patológicos.
- Que se den poco frecuentemente, no quiere decir que podamos desconocerlos e ignorarlos.

Insuficiencia de las Formas Normales

- Con lo visto hasta ahora, se analiza cada relación individualmente siguiendo los principios de:
 - Conservación de los atributos
 - Conservación de la dependencia
- Es necesario especificar todas las DFs en la etapa de diseño.
- Los algoritmos de síntesis requieren calcular el recubrimiento mínimo para el conjunto de DFs. Puede haber mas de un conjunto equivalente y esto puede llevar a diferentes diseños. Algunos con mayor semántica que otros.
- El algoritmo de descomposición es dependiente de la DF que se tome, por lo que puede conducir también a diferentes diseños.

Dependencias Multivaluadas

DMV

- Existen relaciones que poseen restricciones (dependencias) que no pueden ser expresadas mediante DFs .
- Este tipo de dependencia se llama **Dependencia Multivaluada**.
- Es una consecuencia de la 1FN, donde dice que en una tupla, cada atributo puede tener un único valor.
- Esta restricción impide que un atributo posea un conjunto de valores asociados para una tupla.
- Si nos encontramos con la necesidad de almacenar dos o mas atributos multivaluados en una misma relación, deberemos repetir valores de uno de los atributos para cada valor del otro atributo.

Dependencias Multivaluadas

DMV

- Se dice que el atributo A **multidetermina** al atributo B, que se puede escribir $A \twoheadrightarrow B$ cuando sabiendo el valor de A se obtiene un conjunto bien definido de valores de B. O que B es **multidependiente** de A.
- Supongamos que tenemos un conjunto de Materias, Profesores y Libros.
- Cada Materia lo podemos plantear como una jerarquía, cada Materia puede tener distintos Profesores y a su vez pueden utilizar cada uno de los Libros que se utilizan en el mismo.
- Materia determina a un conjunto de Profesores y Materia determina a un conjunto de Libros.
- A su vez Profesores y Libros son independientes entre si.

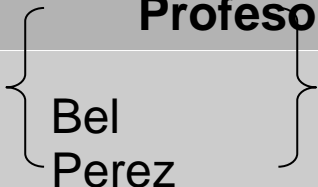
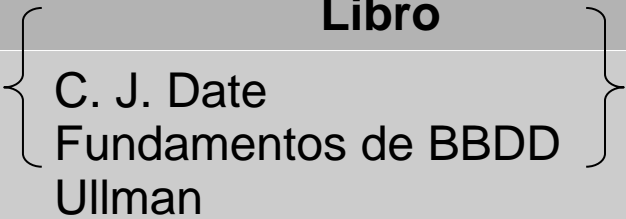
Ejemplo DMV

Materia	Profesor	Libro
Cálculo Diferencial	Agnoli	Piskunov
Cálculo Diferencial	Pintos	Piskunov
Base de Datos.	Bel	C. J. Date
Base de Datos	Bel	Fundamentos de BBDD
Base de Datos	Perez	C. J. Date
Base de Datos	Perez	Fundamentos de BBDD

- La relación MateriaProfesorLibro nos permite almacenar por cada materia que profesores la dictan y que libros utilizan en dicha materia.
- ¿Que sucede si en Base de Datos se incorpora un profesor?
- ¿Qué sucede si en Base de Datos se incorpora un nuevo libro?

Explicación

- A una Materia la dictan un conjunto de Profesores.
- En una Materia utilizan un conjunto de Libros.
- Los atributos Profesor y Libro son independientes entre sí.
- Con esto nos referimos a:
 - Si en una Materia utilizan un Libro. Ese Libro va a ser usado por todos los Profesores que dicten la Materia.
 - Si una Materia es dictada por un Profesor. Ese Profesor va a utilizar todos los libros asociados con la Materia.

Materia	Profesor	Libro
Base de Datos		
Cálculo Diferencial	Agnoli	Piskunov

Primeros Estudios sobre DMV

- El concepto de DMV surge a raíz de redundancias que seguían existiendo en relaciones en 3FN.
- Según Fagin una DMV se escribe de la forma $A \twoheadrightarrow B$, o A multitermina B. O que cierto valor de A implica un conjunto bien definido de valores de B con independencia del resto de los atributos.

Existe DMV:

Si existen las tuplas (m, p_1, t_1) y (m, p_2, t_2) .
Entonces también aparecen las tuplas (m, p_1, t_2) y (m, p_2, t_1) .

Definición de DMV

DMV:

Sea $R\{A, B, C\}$, decimos que $A \twoheadrightarrow B$ si y solo si en todo valor válido posible de R , el conjunto de valores de B que coinciden con un determinado par (a_j, c_j) depende solo de a_j y es independiente de c_j .

- Una DF es un caso particular de DMV, donde para cada valor de A existe un único valor de B y un único valor de C
- Las DMVs se dan de a pares y no de atributos solos

$A \twoheadrightarrow B | C$

Otros Conceptos Relacionados a DMV

- Se cumplen las reglas de:
 - DMV trivial si $A \twoheadrightarrow B|C$ entonces $A \twoheadrightarrow D$ si D son todos los atributos de BC mas algunos de A.
 - Transitividad.
- Pero no se cumple la descomposición.
- Fagin enunció una versión mas sólida del teorema de

Fagin:

Sea $R\{A, B, C\}$, R es igual al join de sus proyecciones $\{A, B\}$ y $\{A, C\}$ si y solo si R satisface La DMV $A \twoheadrightarrow B|C$

Cuarta Forma Normal

- Estos temas estudiados por Fagin, posteriormente se conocieron como la 4FN.

4FN:

La relación R está en 4FN si y solo si siempre que existan subconjuntos A y B de los atributos de R tales que la DMV no trivial $A \twoheadrightarrow B$ se satisfaga, entonces todos los atributos de R son también funcionalmente dependientes de A.

- Una DMV es no trivial cuando ninguno de los atributos de B se encuentra entre los de A.

Analizando el Ejemplo

- MateriaProfesorLibro se encuentra en FNBC, porque no existen DF no triviales que no dependan de la clave primaria (Materia, Profesor, Libro).
- Pero no se encuentra en 4FN porque la DMV $\text{Materia} \rightarrow \rightarrow \text{Profesor}$, no determina funcionalmente a Libro.
- Entonces para normalizar debemos aplicar el Teorema de Fagin sobre esta DMV y nos van a quedar las proyecciones

MateriaProfesor	
Materia	Profesor
Base de Datos	Bel
Base de Datos	Perez
Cálculo Diferencial	Agnoli

MateriaLibro	
Materia	Libro
Base de Datos	C. J. Date
Base de Datos	Ullman
Base de Datos	Fundamentos BBDD
Cálculo Diferencial	Piskunov

Otras Anomalías

- Ya vimos que algunas relaciones que están en FNBC pueden seguir teniendo problemas.
- A continuación veremos que también existen algunas relaciones en 4FN que siguen teniendo problemas.
- Supongamos que tenemos una relación donde almacenamos que libros poseen los alumnos por cada materia.
- Un alumno puede usar uno o mas libros de su preferencia para estudiar una materia.
- Los libros que elija determinado alumno para una materia en particular pueden ser distintos a los que elija otro alumno para la misma materia.
- También se puede dar el caso que un libro sirva para mas de una materia, por lo que el alumno podrá utilizar el libro para todas las materias que sirva.

Ejemplo Relación con Problemas

EstudiaCon		
Alumno	Materia	Libro
Heis	B. Datos	C. J. Date
Heis	B. Datos	Fundamentos BBDD
Gonzales	B. Datos	Ullman
Gonzales	B. Datos Avanzada	Ramakrishnan
Colombo	B. Datos Avanzada	Fundamentos BBDD
Heis	B. Datos Avanzada	Ramakrishnan
Heis	B. Datos Avanzada	Fundamentos BBDD

- ¿Qué sucede si el alumno Heis empieza a cursar B. Datos Avanzada y se compra el libro Ramakrishnan para tal fin?
- También tuvimos que agregar la tupla (Heis, B. Datos Avanzada, Fundamentos BBDD).

Dependencias de Join DJ

- Hasta ahora vimos la forma de descomposición binaria. O sea siempre que hicimos descomposición, obtuvimos 2 relaciones resultantes.
- Una relación puede ser expresada a través del conjunto de todas sus proyecciones válidas posibles.
- Por ejemplo si tenemos la relación:

VAR BASE RELATION persona(dni INT, nombre CHAR(30), apellido CHAR(30), PRIMARY KEY(dni));

- Podríamos expresar a las DJ como:

$*\{\{dni, nombre\}, \{dni, apellido\}\}$

- Una DMV es un caso particular de DJ, donde el número de proyecciones es igual a 2. O sea $r = * (\pi_1(r), \pi_2(r))$

Restricciones de una DJ

- La restricciones que posee la relación EstudiaCon la podemos escribir de la siguiente manera:
 - Si existe $(a1, b1)$ en AB
 - Y existe $(b1, c1)$ en BC
 - Y existe $(a1, c1)$ en AC
 - Entonces existe $(a1, b1, c1)$ en ABC
- También la podemos escribir de la siguiente forma:
 - Si existen las tuplas $(a1, b1, c2)$, $(a2, b1, c1)$, $(a1, b2, c1)$ en ABC
 - Entonces existe $(a1, b1, c1)$

Definición DJ

- En el ejemplo la restricciones nos quedarían:
 - Si existen las tuplas:
 - (Perez, B. Datos Avanzada, Ramakrishnan),
 - (Colombo, B. Datos Avanzada, Fundamentos BBDD),
 - (Perez, B. Datos, Fundamentos BBDD) en EstudiaCon
 - Entonces también existe la tupla (Perez, B. Datos Avanzada, Fundamentos BBDD)

Definición:

Sea R y $\{A, B, C, \dots, Z\}$ subconjuntos de atributos de R . Entonces decimos que R satisface la DJ $\{A, B, C, \dots, Z\}$ si y solo si todo valor válido posible de R es igual al join de sus proyecciones sobre A, B, C, \dots, Z

Analizando el ejemplo

MateriaLibro	
Materia	Libro
B. Datos	C. J. Date
B. Datos	Fundamentos BBDD
B. Datos	Ullman
B. Datos Avanzada	Fundamentos BBDD
B. Datos Avanzada	Ramakrishnan

njoin

AlumnoLibro	
Alumno	Libro
Heis	C. J. Date
Heis	Fundamentos BBDD
Gonzales	Ullman
Gonzales	Ramakrishnan
Colombo	Fundamentos BBDD
Heis	Ramakrishnan

P1		
Alumno	Materia	Libro
Heis	B. Datos	C. J. Date
Heis	B. Datos	Fundamentos BBDD
Colomb o	B. Datos	Fundamentos BBDD
Gonzales	B. Datos	Ullman
Heis	B. Datos Avanzada	Fundamentos BBDD
Colombo	B. Datos Avanzada	Fundamentos BBDD
Gonzales	B. Datos Avanzada	Ramakrishnan
Heis	B. Datos Avanzada	Ramakrishnan

Tupla Espúrea

Analizando el ejemplo continuación

P1		
Alumno	Materia	Libro
Heis	B. Datos	C. J. Date
Heis	B. Datos	Fundamentos BBDD
Colombo	B. Datos	Fundamentos BBDD
Gonzales	B. Datos	Ullman
Heis	B. Datos Avanzada	Fundamentos BBDD
Colombo	B. Datos Avanzada	Fundamentos BBDD
Gonzales	B. Datos Avanzada	
Heis	B. Datos Avanzada	

njoin	
Alumno	Materia
Heis	B. Datos
Heis	B. Datos Avanzada
Gonzales	B. Datos
Gonzales	B. Datos Avanzada
Gonzales	B. Datos Avanzada

Resultado		
Alumno	Materia	Libro
Heis	B. Datos	C. J. Date
Heis	B. Datos	Fundamentos BBDD
Gonzales	B. Datos	Ullman
Heis	B. Datos Avanzada	Fundamentos BBDD
Colombo	B. Datos Avanzada	Fundamentos BBDD
Gonzales	B. Datos Avanzada	Ramakrishnan
Heis	B. Datos Avanzada	Ramakrishnan

Redefinición teorema Fagin

Fagin:

Sea $R\{A, B, C\}$, satisface $DJ^*\{A, B, C\}$ si y solo si satisface la DMV $A \rightarrow \rightarrow BC$

Quinta Forma Normal

Definición:

Una relación R está en 5FN si y solo si cada DJ no trivial válida para R está implicada por las claves candidatas de R

- Una DJ es no trivial si la proyección no son todos los atributos de R .
- Que una DJ esté implicada por las claves candidatas de R significa que la DJ es una superclave de R .

Solución AlumnoMateriaLibro

- Para solucionar las anomalías de la relación AlumnoMateriaLibro es necesario hacer tres proyecciones $\{\text{Alumno}, \text{Materia}\}$, $\{\text{Alumno}, \text{Libro}\}$, $\{\text{Libro}, \text{Materia}\}$.
- Analizando cada una de estas nuevas relaciones:
 - $\{\text{Alumno}, \text{Materia}\}$ se encuentra en 5FN porque cada una de sus DJ de DJ^* es una superclave.
 $DJ^* = \{\{\text{Alumno}, \text{Materia}\}\}$
 - Lo mismo sucede para $\{\text{Alumno}, \text{Libro}\}$ y para $\{\text{Libro}, \text{Materia}\}$

Interpretación 5FN

- Dada la relación:

VAR BASE RELATION persona(dni INT, nombre CHAR(30), apellido CHAR(30), PRIMARY KEY(dni))

- Esta relación se encuentra en FNBC. ¿Comprobar que se encuentra en 5FN?
- La únicas proyecciones de acuerdo a las dependencias sin perdida son $\ast(\{\text{dni}, \text{nombre}\}, \{\text{dni}, \text{apellido}\})$
- Cada una de las DJ está implicada (es una superclave) por la clave primaria DNI. Por lo tanto persona está en 5FN.

Consecuencias 5FN

- La 5FN nos da cuenta que las relaciones se pueden seguir descomponiendo sin tener pérdida.
- Podemos llegar al absurdo de tener una relación diferente por cada par (atributo, clave primaria).
- Pero estas descomposiciones ya no mejoran el diseño.
- Lo que valida que ya no podemos mejorar el diseño descomponiendo las relaciones, es el concepto de que con el join de DJ* obtenemos nuevamente la relación original.
- Por este motivo es que la 5FN se la conoce como la última forma normal. Tiene en cuenta todas las proyecciones posibles que cumplan las dependencias.

Otras Formas Normales

- Pueden existir otras formas normales basadas en otras operaciones del álgebra relacional. Por ejemplo UNION, INTERSECCION, DIFERENCIA, etc.
- El análisis de ellas excede los contenidos de esta materia.
- Además en la práctica son de muy poca utilidad.

Formas Normales Superiores

- Fagín definió una forma normal conocida como Forma Normal de Dominio Clave (FNDK).
- Cualquier relación que se encuentra en FNDK se encuentra en 5FN.
- Esta forma normal en la práctica es muy difícil de llevarla a cabo.
- Al imponer todas las restricciones (DF, DMV, DJ) y además que cualquier valor de atributo pertenezca al dominio, entonces esa relación estará en FNDK

Bibliografía

- Introducción a los Sistemas de Bases de Datos, C. J. Date, Séptima Edición.
- Fundamentos de Sistemas de Bases de Datos, Ramez Elmasri – Shamkant B. Navathe, Tercera Edición.