

Teoría de diseño y normalización



Lineamientos informales de diseño

- Se pueden utilizar como medidas de la calidad del diseño:
 - Asegurar que la semántica de los atributos sea clara en el esquema
 - Reducir la información redundante en las tuplas
 - Reducir los valores NULL en las tuplas
 - Evitar la generación de tuplas espurias

Lineamientos informales de diseño

- Se pueden utilizar como medidas de la calidad del diseño:
 - Asegurar que la semántica de los atributos sea clara en el esquema
 - Reducir la información redundante
 - Reducir los valores NULL en los atributos
 - Evitar la generación de tuplas espurias

Semantica: “significado”,
“sentido”, “interpretación”

1. Semántica clara de los atributos en las relaciones

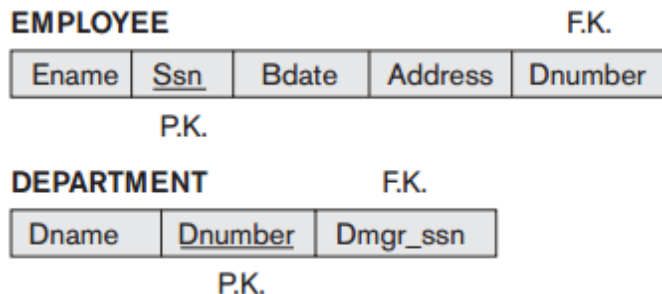
- Cuando se agrupan atributos para formar una relación, normalmente se asume que dichos atributos tienen un significado del mundo real y una interpretación apropiada asociada con ellos.
- La **semántica de una relación** se refiere al significado que se obtiene al interpretar los valores de los atributos de una tupla
- Entre más fácil sea explicar la semántica, mejor es el diseño

1. Semántica clara de los atributos en las relaciones

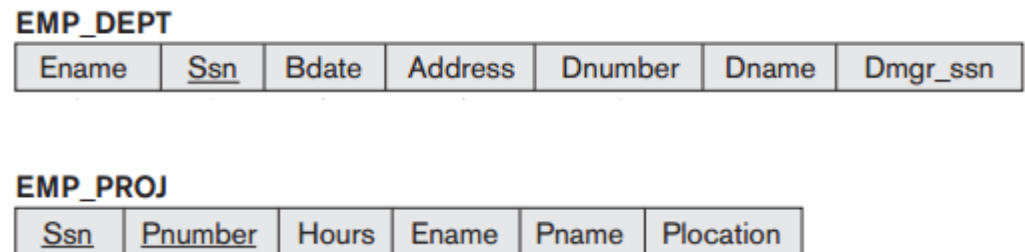
Lineamiento #1

Diseñe un esquema de relación que sea sencillo explicar su significado. No combine atributos de múltiples entidades y asociaciones en una sola relación

Buen diseño



Mal diseño

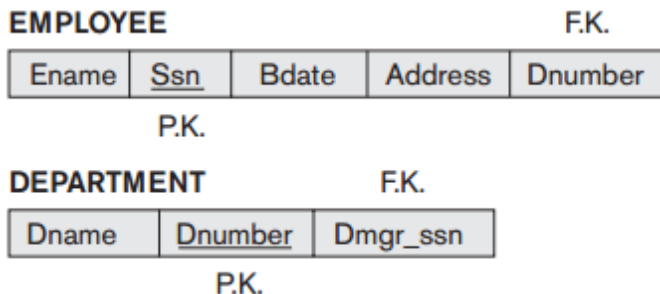


1. Semántica clara de los atributos en las relaciones

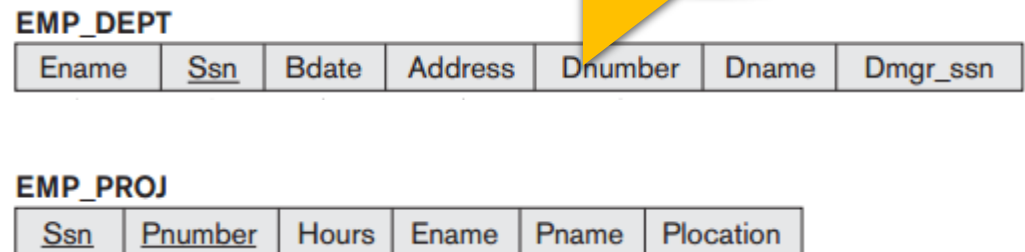
Lineamiento #1

Diseñe un esquema de relación que sea sencillo explicar su significado. No combine atributos de múltiples entidades y asociaciones en una sola relación

Buen diseño



Mal diseño



Mezcla de atributos, es difícil saber de quien es qué

1. Semántica clara de los atributos en las relaciones

Lineamiento #1

Diseñe un esquema de relación que sea sencillo explicar su significado. No combine atributos de multiples entidades y asociaciones en una sola relación

Debería ser una vista, no una relación

Buen diseño

EMPLOYEE

Ename	<u>Ssn</u>	Bdate	Address	Dnumber
-------	------------	-------	---------	---------

P.K.

F.K.

DEPARTMENT

Dname	<u>Dnumber</u>	Dmgr_ssn
-------	----------------	----------

P.K.

F.K.

Mal diseño

EMP_DEPT

Ename	<u>Ssn</u>	Bdate	Address	Dnumber	Dname	Dmgr_ssn
-------	------------	-------	---------	---------	-------	----------

EMP_PROJ

<u>Ssn</u>	<u>Pnumber</u>	Hours	Ename	Pname	Plocation
------------	----------------	-------	-------	-------	-----------

2. Información redundante

					Redundancy	
EMP_DEPT						
Ename	<u>Ssn</u>	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

2. Información redundante

					Redundancy	
EMP_DEPT						
Ename	<u>Ssn</u>	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

Se desperdicia
espacio

2. Información redundante

					Redundancy	
EMP_DEPT						
Ename	<u>Ssn</u>	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

Conduce a
anomalías de
actualización

2. Información redundante

					Redundancy	
EMP_DEPT						
Ename	<u>Ssn</u>	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

Anomalías
de inserción

Conduce a
anomalías de
actualización

Anomalías de
modificación

Anomalías de
eliminación

2. Información redundante

Anomalías de inserción

¿Cómo inserto un nuevo departamento?

EMP_DEPT

Ename	<u>Ssn</u>	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

¿Cómo inserto un nuevo empleado?

2. Información redundante

Anomalías de inserción

EMP_DEPT

					Redundancy	
Ename	<u>Ssn</u>	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

Si se quiere registrar un departamento que aun no tenga empleados asociados, hay que poner NULL a todos los atributos de la tabla de empleados. **Pero la llave primaria Ssn no admite NULL**

2. Información redundante

Anomalías de eliminación

EMP_DEPT

					Redundancy	
Ename	<u>Ssn</u>	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

Si se borran todos los empleados del departamento 5, se perderán los datos del departamento.

2. Información redundante

Anomalías de modificación

EMP_DEPT

					Redundancy	
Ename	<u>Ssn</u>	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

Si se modifica un departamento en una tupla de un empleado, se debe actualizar en las tuplas de todos los empleados del mismo departamento.

2. Información redundante

Lineamiento #2

Diseñe los esquemas de las relaciones de forma tal que no ocurran anomalías de inserción, eliminación y modificación.

Si ya existen anomalías, asegúrese que los programas que actualizan la base de datos, operen adecuadamente.

2. Información redundante

EMP_DEPT					Redundancy	
					Dname	Dmgr_ssn
Ename	<u>Ssn</u>	Bdate	Address	Dnumber		
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

¿Y si necesito hacer esto?...por un tema de performance...

2. Información redundante

Cree una vista o una
vista materializada...

Redundancy

EMP_DEPT

Ename	<u>Ssn</u>	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

2. Información redundante

Cree una vista o una
vista materializada...

EMP_DEPT

Ename	<u>Ssn</u>	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Springfield TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Houston TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Dallas TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Dallas TX	1	Headquarters	888665555

Redundancy

¡Debe utilizar triggers para
mantener actualizadas las
tablas base!

3. Valores NULL

- En ocasiones puede que una relación tenga muchos atributos que no apliquen para todas las tuplas, por lo que se puede terminar en tuplas con muchos valores NULL
- Problemas:
 - Gasto de espacio
 - Problemas de semántica
 - Problemas con los joins
 - Como contarlos al usar funciones agregadas...resultados impredecibles

3. Valores NULL

Lineamiento #3

Evite poner atributos en una relación cuyos valores frecuentemente sean NULL. Si NULL es inevitable asegúrese que solo aplique para casos excepcionales y que no apliquen para la mayoría de las tuplas

Si solo 15% de los empleados tienen oficina, no incluya el atributo oficina en la tabla de empleados. Es mejor mover ese atributo a una tabla aparte

4. Generación de tuplas espurias

- Considere las siguientes relaciones y sus tuplas:

EMP_LOCS

Ename	Plocation
Smith, John B.	Bellaire
Smith, John B.	Sugarland
Narayan, Ramesh K.	Houston
English, Joyce A.	Bellaire
English, Joyce A.	Sugarland
Wong, Franklin T.	Sugarland
Wong, Franklin T.	Houston
Wong, Franklin T.	Stafford
Zelaya, Alicia J.	Stafford
Jabbar, Ahmad V.	Stafford
Wallace, Jennifer S.	Stafford
Wallace, Jennifer S.	Houston
Borg, James E.	Houston

EMP_PROJ1

Ssn	Pnumber	Hours	Pname	Plocation
123456789	1	32.5	ProductX	Bellaire
123456789	2	7.5	ProductY	Sugarland
666884444	3	40.0	ProductZ	Houston
453453453	1	20.0	ProductX	Bellaire
453453453	2	20.0	ProductY	Sugarland
333445555	2	10.0	ProductY	Sugarland
333445555	3	10.0	ProductZ	Houston
333445555	10	10.0	Computerization	Stafford
333445555	20	10.0	Reorganization	Houston
999887777	30	30.0	Newbenefits	Stafford
999887777	10	10.0	Computerization	Stafford
987987987	10	35.0	Computerization	Stafford
987987987	30	5.0	Newbenefits	Stafford
987654321	30	20.0	Newbenefits	Stafford
987654321	20	15.0	Reorganization	Houston
888665555	20	NULL	Reorganization	Houston

4. Generación de tuplas espurias

¿Se puede hacer
join de estas dos
tablas?

- Considere las siguientes relaciones y

EMP_LOCS

Ename	Plocation
Smith, John B.	Bellaire
Smith, John B.	Sugarland
Narayan, Ramesh K.	Houston
English, Joyce A.	Bellaire
English, Joyce A.	Sugarland
Wong, Franklin T.	Sugarland
Wong, Franklin T.	Houston
Wong, Franklin T.	Stafford
Zelaya, Alicia J.	Stafford
Jabbar, Ahmad V.	Stafford
Wallace, Jennifer S.	Stafford
Wallace, Jennifer S.	Houston
Borg, James E.	Houston

EMP_PROJ1

Ssn	Pnumber	Hours	Pname	Plocation
123456789	1	32.5	ProductX	Bellaire
123456789	2	7.5	ProductY	Sugarland
666884444	3	40.0	ProductZ	Houston
453453453	1	20.0	ProductX	Bellaire
453453453	2	20.0	ProductY	Sugarland
333445555	2	10.0	ProductY	Sugarland
333445555	3	10.0	ProductZ	Houston
333445555	10	10.0	Computerization	Stafford
333445555	20	10.0	Reorganization	Houston
999887777	30	30.0	Newbenefits	Stafford
999887777	10	10.0	Computerization	Stafford
987987987	10	35.0	Computerization	Stafford
987987987	30	5.0	Newbenefits	Stafford
987654321	30	20.0	Newbenefits	Stafford
987654321	20	15.0	Reorganization	Houston
888665555	20	NULL	Reorganization	Houston

4. Generación de tuplas espurias

- Considere las siguientes relaciones y

¿Se puede hacer
join de estas dos
tablas?

EMP_LOCS

Ename	Plocation
Smith, John B.	Bellaire
Smith, John B.	Sugarland
Narayan, Ramesh K.	Houston
English, Joyce A.	Bellaire
English, Joyce A.	Sugarland
Wong, Franklin T.	Sugarland
Wong, Franklin T.	Houston
Wong, Franklin T.	Stafford
Zelaya, Alicia J.	Stafford
Jabbar, Ahmad V.	Stafford
Wallace, Jennifer S.	Stafford
Wallace, Jennifer S.	Houston
Borg, James E.	Houston

EMP_PROJ1

Ssn	Pnumber	Hours	Product	Location
123456789	1	32.5	ProductX	Bellaire
123456789	2	7.5	ProductX	Sugarland
666884444	3	40.0	ProductX	Houston
453453453	1	20.0	ProductY	Bellaire
453453453	2	20.0	ProductY	Sugarland
333445555	2	10.0	ProductY	Sugarland
333445555	3	10.0	ProductZ	Houston
333445555	10	10.0	Computerization	Stafford
333445555	20	10.0	Reorganization	Houston
999887777	30	30.0	Newbenefits	Stafford
999887777	10	10.0	Computerization	Stafford
987987987	10	35.0	Computerization	Stafford
987987987	30	5.0	Newbenefits	Stafford
987654321	30	20.0	Newbenefits	Stafford
987654321	20	15.0	Reorganization	Houston
888665555	20	NULL	Reorganization	Houston

Si pero sin
utilizar llaves
primarias

4. Generación de tuplas espurias

- Al hacer un join natural (Plocation) se obtiene lo siguiente:

Ssn	Pnumber	Hours	Pname	Plocation	Ename
123456789	1	32.5	ProductX	Bellaire	Smith, John B.
* 123456789	1	32.5	ProductX	Bellaire	English, Joyce A.
123456789	2	7.5	ProductY	Sugarland	Smith, John B.
* 123456789	2	7.5	ProductY	Sugarland	English, Joyce A.
* 123456789	2	7.5	ProductY	Sugarland	Wong, Franklin T.
666884444	3	40.0	ProductZ	Houston	Narayan, Ramesh K.
* 666884444	3	40.0	ProductZ	Houston	Wong, Franklin T.
* 453453453	1	20.0	ProductX	Bellaire	Smith, John B.
453453453	1	20.0	ProductX	Bellaire	English, Joyce A.
* 453453453	2	20.0	ProductY	Sugarland	Smith, John B.
453453453	2	20.0	ProductY	Sugarland	English, Joyce A.
* 453453453	2	20.0	ProductY	Sugarland	Wong, Franklin T.
* 333445555	2	10.0	ProductY	Sugarland	Smith, John B.
* 333445555	2	10.0	ProductY	Sugarland	English, Joyce A.
333445555	2	10.0	ProductY	Sugarland	Wong, Franklin T.
* 333445555	3	10.0	ProductZ	Houston	Narayan, Ramesh K.
333445555	3	10.0	ProductZ	Houston	Wong, Franklin T.
333445555	10	10.0	Computerization	Stafford	Wong, Franklin T.
* 333445555	20	10.0	Reorganization	Houston	Narayan, Ramesh K.
333445555	20	10.0	Reorganization	Houston	Wong, Franklin T.

*
*
*

4. Generación de tuplas espurias

Si...¿Y? ¿Qué tiene de malo?

- Al hacer un join natural (Plocation) se obtiene lo siguiente

Ssn	Pnumber	Hours	Pname	Plocation	Ename
123456789	1	32.5	ProductX	Bellaire	Smith, John B.
* 123456789	1	32.5	ProductX	Bellaire	English, Joyce A.
123456789	2	7.5	ProductY	Sugarland	Smith, John B.
* 123456789	2	7.5	ProductY	Sugarland	English, Joyce A.
* 123456789	2	7.5	ProductY	Sugarland	Wong, Franklin T.
666884444	3	40.0	ProductZ	Houston	Narayan, Ramesh K.
* 666884444	3	40.0	ProductZ	Houston	Wong, Franklin T.
* 453453453	1	20.0	ProductX	Bellaire	Smith, John B.
453453453	1	20.0	ProductX	Bellaire	English, Joyce A.
* 453453453	2	20.0	ProductY	Sugarland	Smith, John B.
453453453	2	20.0	ProductY	Sugarland	English, Joyce A.
* 453453453	2	20.0	ProductY	Sugarland	Wong, Franklin T.
* 333445555	2	10.0	ProductY	Sugarland	Smith, John B.
* 333445555	2	10.0	ProductY	Sugarland	English, Joyce A.
333445555	2	10.0	ProductY	Sugarland	Wong, Franklin T.
* 333445555	3	10.0	ProductZ	Houston	Narayan, Ramesh K.
333445555	3	10.0	ProductZ	Houston	Wong, Franklin T.
333445555	10	10.0	Computerization	Stafford	Wong, Franklin T.
* 333445555	20	10.0	Reorganization	Houston	Narayan, Ramesh K.
333445555	20	10.0	Reorganization	Houston	Wong, Franklin T.

*
*
*

4. Generación de tuplas espurias

- Al hacer un join natural (Plocation) se obtiene lo siguiente:

Ssn	Pnumber	Hours	Pname	Plocation	Ename
123456789	1	32.5	ProductX	Bellaire	Smith, John B.
123456789	1	32.5	ProductX	Bellaire	English, Joyce A.
123456789	2	7.5	ProductY	Sugarland	Smith, John B.
123456789	1	32.5	ProductX	Bellaire	English, Joyce A.
666888444	3	40.0	ProductZ	Houston	Narayan, Ramesh K.
666888444	3	40.0	ProductZ	Houston	Wong, Franklin T.
333445555	1	20.0	ProductX	Bellaire	Smith, John B.
333445555	1	20.0	ProductX	Bellaire	English, Joyce A.
333445555	2	20.0	ProductY	Sugarland	Smith, John B.
333445555	2	20.0	ProductY	Sugarland	English, Joyce A.
333445555	2	20.0	ProductY	Sugarland	Wong, Franklin T.
333445555	2	10.0	ProductY	Sugarland	Smith, John B.
333445555	2	10.0	ProductY	Sugarland	English, Joyce A.
333445555	2	10.0	ProductY	Sugarland	Wong, Franklin T.
333445555	3	10.0	ProductZ	Houston	Narayan, Ramesh K.
333445555	3	10.0	ProductZ	Houston	Wong, Franklin T.
333445555	10	10.0	Computerization	Stafford	Wong, Franklin T.
333445555	20	10.0	Reorganization	Houston	Narayan, Ramesh K.
333445555	20	10.0	Reorganization	Houston	Wong, Franklin T.

Esta tupla no tiene sentido!

4. Generación de tuplas espurias

- Al hacer un join natural (Plocation) se obtiene lo siguiente:

Ssn	Pnumber	Hours	Pname	Plocation	Ename
123456789	1	32.5	ProductX	Bellaire	Smith, John B.
* 123456789	1	32.5	ProductX	Bellaire	English, Joyce A.
123456789	2	7.5	ProductY	Sugarland	Smith, John B.

123456789	1	32.5	ProductX	Bellaire	English, Joyce A.
-----------	---	------	----------	----------	-------------------

666884444	3	40.0	ProductZ	Houston	Narayan, Ramesh K.
* 666884444	3	40.0	ProductZ	Houston	Wong, Franklin T.
* 453453453	1	20.0	ProductX	Bellaire	Smith, John B.
453453453	1	20.0	ProductX	Bellaire	English, Joyce A.
* 453453453	2	20.0	ProductY	Sugarland	Smith, John B.

EMP_PROJ

<u>Ssn</u>	<u>Pnumber</u>	Hours	Ename	Pname	Plocation
123456789	1	32.5	Smith, John B.	ProductX	Bellaire
123456789	2	7.5	Smith, John B.	ProductY	Sugarland
666884444	3	40.0	Narayan, Ramesh K.	ProductZ	Houston
453453453	1	20.0	English, Joyce A.	ProductX	Bellaire
453453453	2	20.0	English, Joyce A.	ProductY	Sugarland
333445555	2	10.0	Wong, Franklin T.	ProductY	Sugarland
333445555	3	10.0	Wong, Franklin T.	ProductZ	Houston

4. Generación de tuplas espurias

Lineamiento #4

Diseñe relaciones que puedan unirse con condiciones de igualdad en atributos apropiadamente relacionados (llave primaria – llave foránea) garantizando que no se genere

En resumen...

- Los lineamientos vistos aunque son informales, permiten determinar el nivel de calidad del diseño
- Las anomalías deben evitarse al máximo
- Los NULL deben ser cuidadosamente utilizados únicamente en situaciones que a fuerza los requieran
- Se deben relacionar relaciones de forma llave primaria – llave foránea para evitar tuplas espurias
- Ahora bien...**¿Cómo se definen estos lineamientos más formalmente?**

Dependencias funcionales

- Una dependencia funcional es una propiedad del esquema de la relación
- Los diseñadores determinan las dependencias funcionales a partir de la semántica. Por ejemplo, si se está diseñado una base de datos de persona estadounidenses, a partir de la semántica se deriva la siguiente dependencia:

$$\{\text{State, Driver_license_number}\} \rightarrow \text{Ssn}$$

- Otros ejemplos de dependencias funcionales (tabla EMP_PROJ):

$$\text{Ssn} \rightarrow \text{Ename}$$

$$\text{Pnumber} \rightarrow \{\text{Pname, Plocation}\}$$

$$\{\text{Ssn, Pnumber}\} \rightarrow \text{Hours}$$

Dependencias funcionales

- Una dependencia funcional es una propiedad del esquema de la relación

- Los diseñadores determinan las dependencias funcionales a partir de la semántica. Por ejemplo, si se está considerando una relación de personas estadounidenses, a partir de la siguiente dependencia:

$\{ \text{State, D} \}$

Dado un valor de Ssn,
se determina el
Ename

La misma cédula no
puede estar asociada a
dos nombres
diferentes...

- Otros ejemplos de dependencias funcionales (para la relación EMP_PROJ):

$\text{Ssn} \rightarrow \text{Ename}$

$\text{Pnumber} \rightarrow \{ \text{Pname, Plocation} \}$

$\{ \text{Ssn, Pnumber} \} \rightarrow \text{Hours}$

Dependencias funcionales

- Otro ejemplo:

TEACH

Teacher	Course	Text
Smith	Data Structures	Bartram
Smith	Data Management	Martin
Hall	Compilers	Hoffman
Brown	Data Structures	Horowitz

- Se cumple que:

TEXT → COURSE

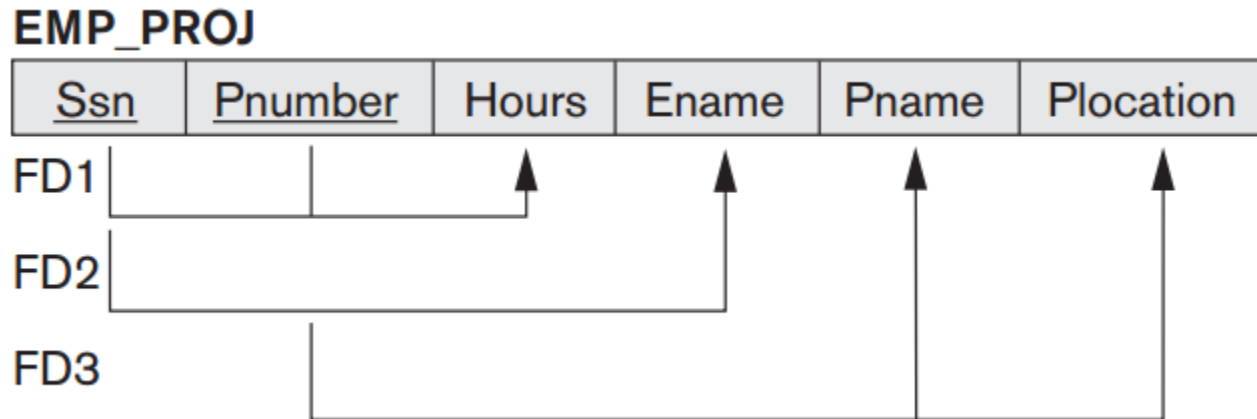
- Pero no existe la dependencia:

TEACHER → COURSE

¿Por qué?

Dependencias funcionales

- ¿Cómo se representa una dependencia funcional en un diagrama?



Formas normales

- El proceso de normalización fue definido por Codd en 1972
- Toma una relación y la “pasa” por una serie de pruebas para certificar que satisface cierta forma normal. Es un proceso top-down que toma cada relación y la descompone para alcanzar formas más altas
- Existen 5 formas normales (1FN, 2FN, 3FN, 4FN y 5FN)

Forma normal se refiere a la condición más alta que alcance y define el grado de normalización

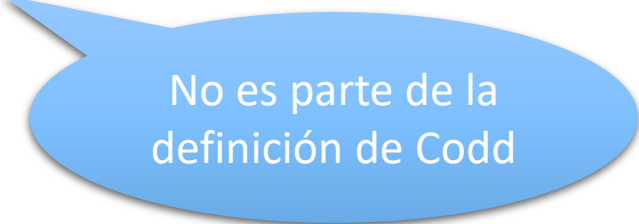
Desnormalización es el proceso de unir relaciones y almacenarlas como tablas base en una forma inferior.

Primera forma normal (1FN)

- Es parte de la definición formal de una relación
- Establece que el dominio de un atributo debe incluir únicamente valores atómicos y que el valor de cualquier atributo debe ser single-valued
- Evita que se formen relaciones dentro de relaciones o relaciones como atributos de otras
- Establece que toda relación debe tener una llave primaria (puede ser compuesta)

Primera forma normal (1FN)

- Es parte de la definición formal de una relación
- Establece que el dominio de un atributo debe incluir únicamente valores atómicos y que el valor de cualquier atributo debe ser single-valued
- Evita que se formen relaciones dentro de relaciones o relaciones como atributos de otras
- Establece que toda relación debe tener una llave primaria (puede ser compuesta)



No es parte de la
definición de Codd

Primera forma normal (1FN)

DEPARTMENT

Dname	<u>Dnumber</u>	Dmgr_ssn	Dlocations
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

Primera forma normal (1FN)

DEPARTMENT

Dname	<u>Dnumber</u>	Dmgr_ssn	Dlocation
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

No está en
1FN

¿Cómo normalizarla?

Primera forma normal (1FN)

DEPARTMENT

Dname	<u>Dnumber</u>	Dmgr_ssn	Dlocations
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

#2 Expandir la llave de DEPARTMENT para incluir cada uno de Dlocations. Esta solución introduce redundancia.

Primera forma normal (1FN)

DEPARTMENT

DEPARTMENT				
Res	Dname	<u>Dnumber</u>	Dmgr_ssn	<u>Dlocation</u>
Adn	Research	5	333445555	Bellaire
Hea	Research	5	333445555	Sugarland
	Research	5	333445555	Houston
	Administration	4	987654321	Stafford
	Headquarters	1	888665555	Houston

#2 Expandir la llave de DEPARTMENT para incluir cada uno de Dlocations. Esta solución introduce redundancia.

Primera forma normal (1FN)

DEPARTMENT

Dname	<u>Dnumber</u>	Dmgr_ssn	Dlocations
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

#3 Crear una columna para cada valor de Dlocations. Esto puede hacerse solo si la cantidad de valores de Dlocations es poca. Tiene el problema que muchas de estas nuevas columnas tendrán el valor en NULL.

Primera forma normal (1FN)

DEPARTMENT

Dname	<u>Dnumber</u>	Dmgr_ssn	Dlocations
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

#1 Remover el atributo que viola la 1FN y ponerlo en una tabla separada junto con la llave primaria Dnumber de DEPARTMENT. La llave de esta nueva tabla sería {Dnumber, DLocation}. **Descompone una relación no-1FN en dos 1FN**

Primera forma normal (1FN)

DEPARTMENT

Dname	<u>Dnumber</u>	Dmgr_ssn	Dlocations
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

Forma
correcta!

#1 Remover el atributo que viola la 1FN y ponerlo en una tabla separada junto con la llave primaria Dnumber de DEPARTMENT. La llave de esta nueva tabla sería {Dnumber, DLocation}. **Descompone una relación no-1FN en dos 1FN**

Segunda forma normal (2FN)

- Se basa en el concepto de dependencia funcional completa.
- Una dependencia funcional $X \rightarrow Y$ es completa si remover cualquier atributo de X significa que la dependencia no se mantiene
- Una dependencia funcional $X \rightarrow Y$ es parcial si algunos atributos de X se pueden remover y la dependencia aún se mantendrá

Segunda forma normal (2FN)

- Se basa en el concepto de dependencia funcional completa.
- Una dependencia funcional $X \rightarrow Y$ es completa si remover cualquier atributo de X significa que la dependencia no se mantiene
- Una dependencia funcional $X \rightarrow Y$ es parcial si algunos atributos de X se pueden remover y la dependencia aún se mantendrá

$$\{\text{Ssn}, \text{Pnumber}\} \rightarrow \text{Hours}$$

Segunda forma normal (2FN)

- Se basa en el concepto de dependencia funcional completa.
- Una dependencia funcional $X \rightarrow Y$ es completa si remover cualquier atributo de X significa que la dependencia no se mantiene
- Una dependencia funcional $X \rightarrow Y$ es parcial si algunos atributos de X se pueden remover y la dependencia aún se mantendrá

Es una dependencia
funcional completa

$\{\text{Ssn, Pnumber}\} \rightarrow \text{Hours}$

Segunda forma normal (2FN)

- Se basa en el concepto de dependencia funcional completa.
- Una dependencia funcional $X \rightarrow Y$ es completa si remover cualquier atributo de X significa que la dependencia no se mantiene
- Una dependencia funcional $X \rightarrow Y$ es parcial si algunos atributos de X se pueden remover y la dependencia aún se mantiene



¿Por qué?

$\{Ssn, Pnumber\} \rightarrow Hours$

Segunda forma normal (2FN)

- Se basa en el concepto de dependencia funcional completa.
- Una dependencia funcional $X \rightarrow Y$ es completa si remover cualquier atributo de X significa que la dependencia no se mantiene
- Una dependencia funcional $X \rightarrow Y$ es parcial si se pueden remover atributos de X y la dependencia aún se mantiene

Un mismo empleado puede trabajar la misma cantidad de horas, en diferentes proyectos

$\{\text{Ssn}, \text{Pnumber}\} \rightarrow \text{Hours}$

Segunda forma normal (2FN)

- Se basa en el concepto de dependencia funcional completa.
- Una dependencia funcional $X \rightarrow Y$ es completa si remover cualquier atributo de X significa que la dependencia no se mantiene
- Una dependencia funcional $X \rightarrow Y$ es parcial si se pueden remover atributos de X y la dependencia aún se mantiene

Hours no depende funcionalmente solo de Ssn

$\{\text{Ssn}, \text{Pnumber}\} \rightarrow \text{Hours}$

Segunda forma normal (2FN)

- Se basa en el concepto de dependencia funcional completa.
- Una dependencia funcional $X \rightarrow Y$ es completa si remover cualquier atributo de X significa que la dependencia no se mantiene
- Una dependencia funcional $X \rightarrow Y$ es completa si X se pueden remover y la dependencia no se mantiene

Hours si depende funcionalmente de { Ssn, Pnumber} puesto que para un proyecto dado, un empleado solo puede trabajar x Hours

$\{Ssn, Pnumber\} \rightarrow Hours$

Segunda forma normal (2FN)

- Se basa en el concepto de dependencia funcional completa.
- Una dependencia funcional $X \rightarrow Y$ es completa si remover cualquier atributo de X significa que la dependencia no se mantiene
- Una dependencia funcional $X \rightarrow Y$ es parcial si algunos atributos de X se pueden remover y la dependencia

No hay dos registros
con el mismo {Ssn,
Pnumber}

$\{\text{Ssn}, \text{Pnumber}\} \rightarrow \text{Hours}$

Segunda forma normal (2FN)

- Se basa en el concepto de dependencia funcional completa.
- Una dependencia funcional $X \rightarrow Y$ es completa si remover cualquier atributo de X significa que la dependencia no se mantiene
- Una dependencia funcional $X \rightarrow Y$ es parcial si algunos atributos de X se pueden remover y la dependencia



Es una dependencia
parcial

$\{\text{Ssn}, \text{Pnumber}\} \rightarrow \text{Ename}$

Segunda forma normal (2FN)

- Se basa en el concepto de dependencia funcional completa.
- Una dependencia funcional $X \rightarrow Y$ es completa si remover cualquier atributo de X significa que la dependencia no se mantiene
- Una dependencia funcional $X \rightarrow Y$ es parcial si algunos atributos de X se pueden remover y la dependencia se mantendrá

Aún si quito Pnumber, Ename
sigue dependiendo de Ssn

$\{\text{Ssn}, \text{Pnumber}\} \rightarrow \text{Ename}$

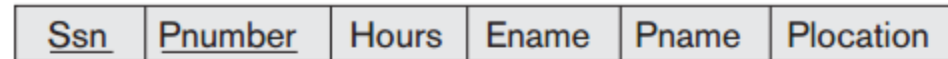
Segunda forma normal (2FN)

- Entonces, una relación está en 2FN si:
Cada atributo no llave es funcionalmente dependiente de la llave primaria
- Para verificar que una relación esté en 2FN se toma cada atributo que no sea llave (o llave candidata) y se verifica que sean funcionalmente dependiente de la llave completa
- Si la llave no es compuesta, es más fácil verificar que estén en 2FN, pero esto no implica que utilizar llaves compuestas sea malo
- ¿Cómo se normaliza una relación que esté en 1FN pero no en 2FN?

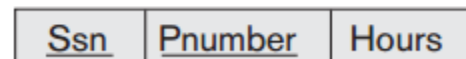
Segunda forma normal (2FN)

- Entonces, una relación está en 2FN si:
Cada atributo no llave es funcionalmente dependiente de la llave primaria
- Para verificar que una relación esté en 2FN se toma cada atributo que no sea llave (o llave candidata) y se verifica que sean funcionalmente dependiente de la llave completa
- Si la llave no es compuesta, es más fácil verificar que estén en 2FN, pero esto no implica que utilizar llaves compuestas sea malo
- ¿Cómo se normaliza una relación que esté en 1FN pero no en 2FN?

Segunda forma normal (2FN)



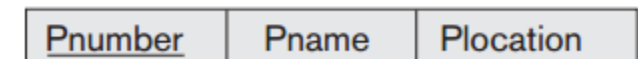
2NF Normalization



EP2

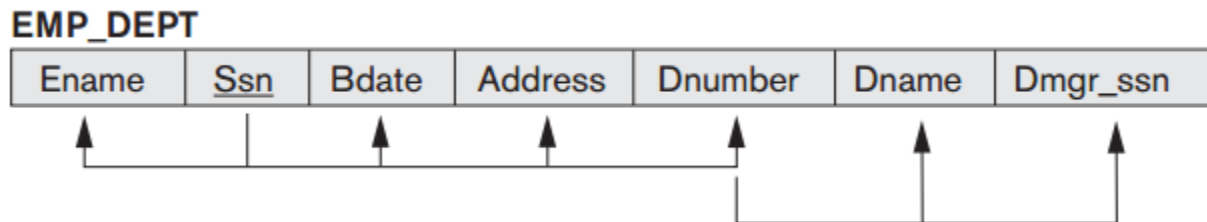


EP3



Tercera forma normal (3FN)

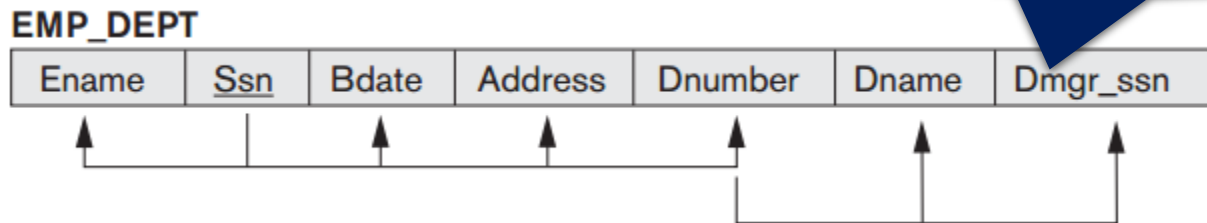
- La tercera forma normal se basa en el concepto de dependencia transitiva.
- Una dependencia funcional $X \rightarrow Y$ es transitiva si existe un conjunto de atributos Z en R que no son ni llave candidata o parte de la llave primaria de R y se cumple que $X \rightarrow Z$ y $Z \rightarrow Y$
- Una relación R está en 3FN si está en 2FN y ningún atributo no-primario es dependiente transitivamente de la llave primaria



Tercera forma normal (3FN)

- La tercera forma normal se basa en el concepto de dependencia transitiva.
- Una dependencia funcional $X \rightarrow Y$ es transitiva si existe un conjunto de atributos Z en R que no son ni llave candidata o parte de la llave primaria de R y se cumple que $X \rightarrow Z$ y $Z \rightarrow Y$
- Una relación R está en 3FN si está en 2FN y ningún atributo no llave primario es dependiente transitivamente

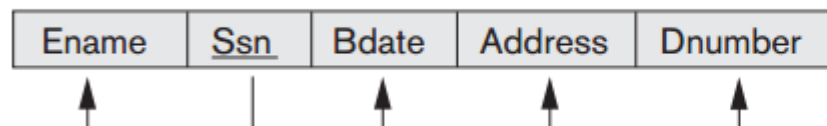
¿Cómo convertirla a 3FN?



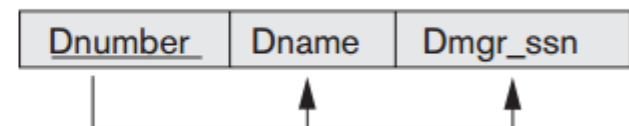
Tercera forma normal (3FN)

- La tercera forma normal se basa en el concepto de dependencia transitiva.
- Una dependencia funcional $X \rightarrow Y$ es transitiva si existe un conjunto de atributos Z en R que no son ni llave candidata o parte de la llave primaria de R y se cumple que $X \rightarrow Z$ y $Z \rightarrow Y$
- Una relación R está en 3FN si está en 2FN y ningún atributo no-primario es dependiente transitivamente de la llave primaria

ED1



ED2

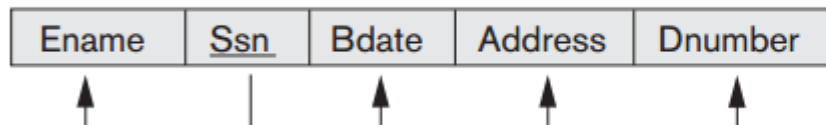


Resumen 1FN – 3FN (3FN)

- La tercera forma normal se basa en el concepto de dependencia transitiva.
- Una dependencia funcional $X \rightarrow Y$ es transitiva si existe un conjunto de atributos Z en R que no son ni llave candidata o parte de la llave primaria de R y se cumple que $X \rightarrow Z$ y $Z \rightarrow Y$
- Una relación R está en 3FN si está en 2FN y ningún atributo no llave candidato es dependiente transitivamente de otro atributo no llave candidato

Nuevamente, se aplicó descomposición

ED1



ED2



Tercera forma normal (3FN)

Table 15.1 Summary of Normal Forms Based on Primary Keys and Corresponding Normalization

Normal Form	Test	Remedy (Normalization)
First (1NF)	Relation should have no multivalued attributes or nested relations.	Form new relations for each multivalued attribute or nested relation.
Second (2NF)	For relations where primary key contains multiple attributes, no nonkey attribute should be functionally dependent on a part of the primary key.	Decompose and set up a new relation for each partial key with its dependent attribute(s). Make sure to keep a relation with the original primary key and any attributes that are fully functionally dependent on it.
Third (3NF)	Relation should not have a nonkey attribute functionally determined by another nonkey attribute (or by a set of nonkey attributes). That is, there should be no transitive dependency of a nonkey attribute on the primary key.	Decompose and set up a relation that includes the nonkey attribute(s) that functionally determine(s) other nonkey attribute(s).

Cuarta forma normal (4FN)

- De acuerdo a lo que se establece en 1FN, un atributo de una relación no puede tener un conjunto de valores, es decir no puede ser multivaluado.
- Vimos varias estrategias para convertir una relación con atributos multivaluados a una relación en 1FN, una de las cuales era expandir la llave para incluir cada uno de los valores del atributo multivaluado:


DEPARTMENT

Dname	<u>Dnumber</u>	Dmgr_ssn	Dlocations
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

Cuarta forma normal (4FN)

- De acuerdo a lo que se establece en 1FN, un atributo de una relación no puede tener un conjunto de valores, es decir no puede ser multivaluado.
- Vimos varias estrategias para convertir una relación con atributos multivaluados a una relación en 1FN, una de las cuales era expandir la llave para incluir cada uno de los valores del atributo multivaluado:

DEPARTMENT



Dname	<u>Dnumber</u>	Dmgr_ssn	<u>Dlocation</u>
Research	5	333445555	Bellaire
Research	5	333445555	Sugarland
Research	5	333445555	Houston
Administration	4	987654321	Stafford
Headquarters	1	888665555	Houston

Cuarta forma normal (4FN)

- Si la relación tuviera dos o más atributos multivaluados no correlacionados que se normalizaron con la estrategia anterior, se tendría una dependencia multivaluada
- Por ejemplo:

<u>Ename</u>	<u>Pname</u>	<u>Dname</u>
Smith	X	John
Smith	Y	Anna
Smith	X	Anna
Smith	Y	John

Cuarta forma normal (4FN)

- Si la relación tuviera dos o más atributos multivaluados no correlacionados que se normalizaron con la estrategia anterior, se tendría una dependencia multivaluada

- Por ejemplo:

Nombre de
proyecto

Nombre de
dependiente

<u>Ename</u>	<u>Pname</u>	<u>Dname</u>
Smith	X	John
Smith	Y	Anna
Smith	X	Anna
Smith	Y	John

Cuarta forma normal (4FN)

- Si la relación tuviera dos o más atributos multivaluados no correlacionados que se normalizaron con la estrategia anterior, se tendría una dependencia multivaluada

**Pname y Dname no
están correlacionadas...**

- Por ejemplo:

<u>Ename</u>	<u>Pname</u>	<u>Dname</u>
Smith	X	John
Smith	Y	Anna
Smith	X	Anna
Smith	Y	John

Cuarta forma normal (4FN)

- Si la relación tuviera dos o más atributos multivaluados no correlacionados que se normalizaron con la estrategia anterior, se tendría una dependencia

**Ambas son
multivaluadas**

- Por ejemplo:

<u>Ename</u>	<u>Pname</u>	<u>Dname</u>
Smith	X	John
Smith	Y	Anna
Smith	X	Anna
Smith	Y	John

Cuarta forma normal (4FN)

- Si la relación tuviera dos o más atributos multivaluados no correlacionados que se normalizaron con la estrategia anterior, se tendría una dependencia multivaluada
- Por ejemplo:

<u>Ename</u>	<u>Pname</u>	<u>Dname</u>
Smith	X	John
Smith	Y	Anna
Smith	X	Anna
Smith	Y	John

Para tener todos los valores de Dname hay que repetir Ename y Pname o usar NULL en Pname

Cuarta forma normal (4FN)

Para tener todos los valores de Pname hay que repetir Ename y Dname o usar NULL en Dname

Si dos o más atributos multivaluados no se normalizaron con la estrategia anterior, se multivaluada

<u>Ename</u>	<u>Pname</u>	<u>Dname</u>
Smith	X	John
Smith	Y	Anna
Smith	X	Anna
Smith	Y	John

Cuarta forma normal (4FN)

- Si la relación tuviera dos o más atributos multivaluados no correlacionados que se normalizaron con la estrategia anterior, se tendría una dependencia multivaluada
- Por ejemplo:

<u>Ename</u>	<u>Pname</u>	<u>Dname</u>
Smith	X	John
Smith	Y	Anna
Smith	X	Anna
Smith	Y	John

¿Cómo se soluciona?

Cuarta forma normal (4FN)

- Si la relación tuviera dos o más atributos multivaluados no correlacionados que se normalizaron con la estrategia anterior, se tendría una dependencia multivaluada
- Por ejemplo:

EMP_PROJECTS

<u>Ename</u>	<u>Pname</u>
Smith	X
Smith	Y

EMP_DEPENDENTS

<u>Ename</u>	<u>Dname</u>
Smith	John
Smith	Anna

Quinta forma normal (5FN)

- Dependencia de reunión.
- Se presentan muy rara vez y es difícil detectarlos en la práctica.

SUPPLY

<u>Sname</u>	<u>Part_name</u>	<u>Proj_name</u>
Smith	Bolt	ProjX
Smith	Nut	ProjY
Adamsky	Bolt	ProjY
Walton	Nut	ProjZ
Adamsky	Nail	ProjX
Adamsky	Bolt	ProjX
Smith	Bolt	ProjY

Quinta forma normal (5FN)

SUPPLY

<u>Sname</u>	<u>Part name</u>	<u>Proj name</u>
Smith	Bolt	ProjX
Smith	Nut	ProjY
Adamsky	Bolt	ProjY
Walton	Nut	ProjZ
Adamsky	Nail	ProjX
Adamsky	Bolt	ProjX
Smith	Bolt	ProjY

R_1

<u>Sname</u>	<u>Part name</u>
Smith	Bolt
Smith	Nut
Adamsky	Bolt
Walton	Nut
Adamsky	Nail

R_2

<u>Sname</u>	<u>Proj name</u>
Smith	ProjX
Smith	ProjY
Adamsky	ProjY
Walton	ProjZ
Adamsky	ProjX

R_3

<u>Part name</u>	<u>Proj name</u>
Bolt	ProjX
Nut	ProjY
Bolt	ProjY
Nut	ProjZ
Nail	ProjX

Teoría de diseño y normalización

