

# Normalización

Base de Datos - Informática II  
UPA

Prof. Jorge Meza & Prof. Carlos Núñez

# Quote

*“In God we trust; all other must bring data.”.*

*Willian Edwards Deming.*

# Diseño relacional

Identificar Entidades

Identificar Relaciones

Transformar

*Cómo sabemos cual modelado es mejor que otro?*

# Objetivo del modelado?

- Relaciones
  - Preservación de la información
  - Redundancia mínima

# Guías informales para el modelado

- Asegurarse que la semántica de los atributos es clara
- Reducir la información redundante
- Reducir los valores NULL
- Eliminar la posibilidad de tablas espurias

# Guías informales para el modelado

Semántica clara

EMPLOYEE					F.K.
Ename	<u>Ssn</u>	Bdate	Address	Dnumber	

P.K.

DEPARTMENT			F.K.
Dname	<u>Dnumber</u>	Dmgr_ssn	

P.K.

DEPT_LOCATIONS		F.K.
<u>Dnumber</u>	<u>Dlocation</u>	

P.K.

PROJECT				F.K.
Pname	<u>Pnumber</u>	Plocation	Dnum	

P.K.

WORKS_ON			F.K.	F.K.
<u>Ssn</u>	<u>Pnumber</u>	Hours		

P.K.

# Guía #1

- Diseñar de forma que sea fácil explicar
- No combinar  
entidades/hechos/relaciones en una sola  
entidad

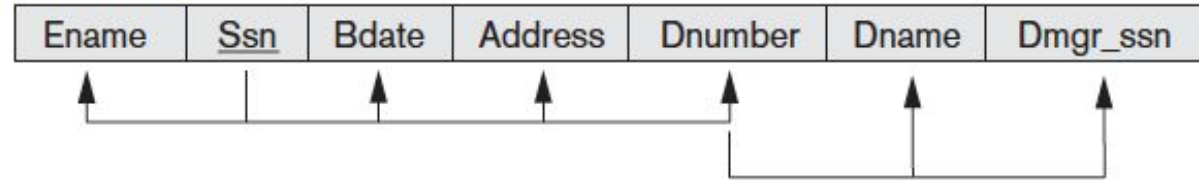
# Guía #1 - Ejemplo de no cumplimiento

**Figure 15.3**

Two relation schemas suffering from update anomalies. (a) EMP\_DEPT and (b) EMP\_PROJ.

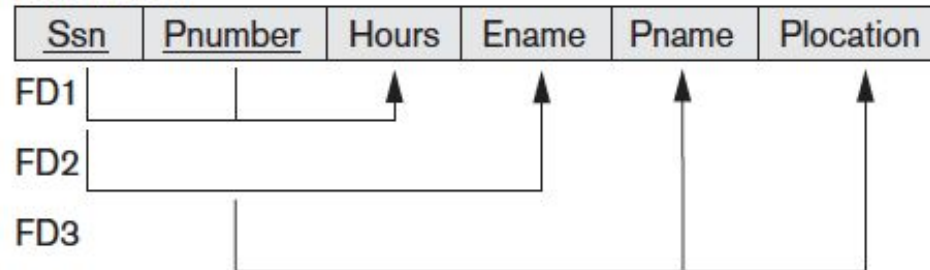
(a)

EMP\_DEPT



(b)

EMP\_PROJ



*Información de entidades combinadas!*



# Guía #1 - Anomalías

- Almacenamiento
- Anomalías de inserción (insert)
- Anomalías de actualización (update)
- Anomalías de eliminación (Deletion)

# Guía #1 - Anomalías

EMP\_DEPT

Ename	Ssn	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	1				
English, Joyce A.	453453453	1				
Jabbar, Ahmad V.	987987987	1				
Borg, James E.	888665555	1				

Redundancy

Redundancy

Redundancy

EMP\_PROJ

Ssn	Pnumber	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
333445555	10	10.0	Wong, Franklin T.	Computerization	Stafford

- Redundancia
- Insertar un nuevo empleado?
- Insertar un nuevo departamento?
- Borrar uno?

## Guía #2

- Diseñar de forma que las anomalías no existan



## Guía #3

- Diseñar de forma que los valores NULL sean poco frecuentes
- *Ejemplo: Si sólo el 15% de empleados tiene oficinas individuales, entonces no justifica incluir un atributo "nro\_oficina" en la entidad empleado. En vez de ello, se puede tener una entidad "OFICINA\_EMP" para incluir solo a aquellos que tengan despachos individuales.*

# Diseño relacional

*Cómo sabemos que tan **bueno** o **malo** es  
nuestro modelado?*

# Dependencia Funcional

$X \rightarrow Y$ , donde  $X$  e  $Y$  son conjuntos de atributos. Esto significa que, los valores de  $Y$  dependen (o *son determinados por*), los valores de  $X$ ; Por otro lado, los valores de  $X$  de una fila *determinan unívocamente* a los valores de  $Y$ .

- *Existe una dependencia funcional de  $X$  a  $Y$*
- *$Y$  es funcionalmente dependiente de  $X$*

a.  $Ssn \rightarrow Ename$

b.  $Pnumber \rightarrow \{Pname, Plocation\}$

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

*DF Compuesta o Completa*

# Diseño relacional "Por Análisis". Formas Normales

Toma una relación y la pasa a través de algunos "tests" para verificar que satisface una "forma normal".

# 1FN (primera Forma Normal)

- Rechazar atributos multivalores
- Rechazar atributos compuestos
- Cualquier combinación de los anteriores
- Debe incluir valores atómicos (simples, indivisibles)
- *No existen registros duplicados \**
- *Debe existir una independencia del orden tanto de las filas como de las columnas, es decir, si los datos cambian de orden no deben cambiar sus significados \**




# 1NF

(a)

DEPARTMENT

Dname	<u>Dnumber</u>	Dmgr_ssn	Dlocations



(b)

DEPARTMENT

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

(c)

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

*Elimino los multivalores, pero quedo con redundancia!*

# 1NF

**Customer**

Customer ID	First Name	Surname	Telephone Number
123	Pooja	Singh	555-861-2025, 192-122-1111
456	Zhang	San	(555) 403-1659 Ext. 53; 182-929-2929
789	John	Doe	555-808-9633

**Customer**

Customer ID	First Name	Surname	Telephone Number
123	Pooja	Singh	555-861-2025
123	Pooja	Singh	192-122-1111
456	Zhang	San	182-929-2929
456	Zhang	San	(555) 403-1659 Ext. 53
789	John	Doe	555-808-9633

**Customer Name**

<u>Customer ID</u>	First Name	Surname
123	Pooja	Singh
456	Zhang	San
789	John	Doe

**Customer Telephone Number**

<u>Customer ID</u>	<u>Telephone Number</u>
123	555-861-2025
123	192-122-1111
456	(555) 403-1659 Ext. 53
456	182-929-2929
789	555-808-9633

*Elimino la  
redundancia!*

## 2FN (Segunda Forma Normal)

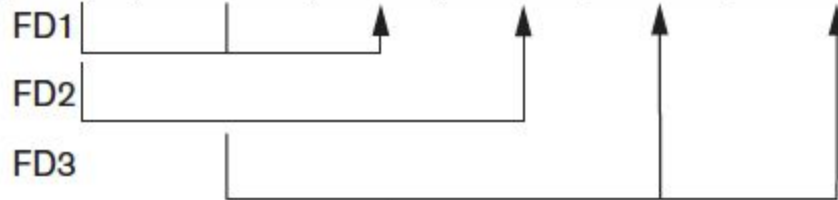
- Una entidad está en segunda forma normal, si ella está en la 1FN y sus atributos no clave son funcionalmente dependientes de su clave completa
  - *Si existe un solo atributo clave, entonces 1FN ~ 2FN*

# 2FN (Segunda Forma Normal)

(a)

EMP\_PROJ

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



2NF Normalization

EP1

<u>Ssn</u>	<u>Pnumber</u>	Hours
------------	----------------	-------



EP2

<u>Ssn</u>	Ename
------------	-------



EP3

<u>Pnumber</u>	Pname	Plocation
----------------	-------	-----------



# Dependencia Transitiva

- Dados los atributos "A", "B" y "C", siendo "A" la clave primaria, se dice que "B", "C" son dependientes transitivos si y solamente si, fueran funcionalmente dependientes de "A", además de existir una dependencia funcional entre B y C.

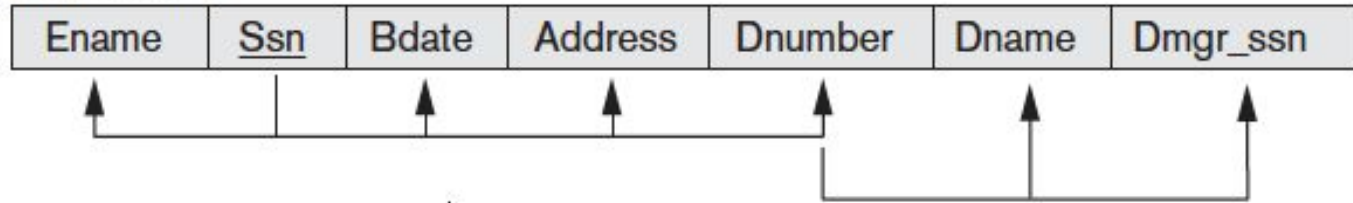
## 3FN (Tercera Forma Normal)

- Una entidad está en tercera forma normal, si ella está en la 2FN y no posee dependencias transitivas, es decir no hay dependencia de atributos no clave entre sí.

# 3FN (Tercera Forma Normal)

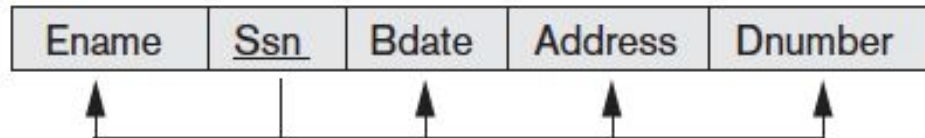
(b)

EMP\_DEPT

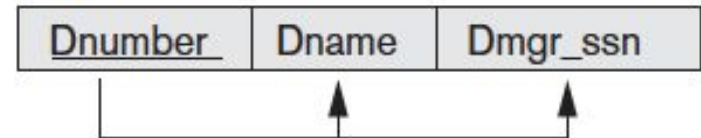


3NF Normalization

ED1



ED2



# 3FN (Tercera Forma Normal)

**Tournament Winners**

<u>Tournament</u>	<u>Year</u>	<u>Winner</u>	<u>Winner Date of Birth</u>
Indiana Invitational	1998	Al Fredrickson	21 July 1975
Cleveland Open	1999	Bob Albertson	28 September 1968
Des Moines Masters	1999	Al Fredrickson	21 July 1975
Indiana Invitational	1999	Chip Masterson	14 March 1977

**Tournament Winners**

<u>Tournament</u>	<u>Year</u>	<u>Winner</u>
Indiana Invitational	1998	Al Fredrickson
Cleveland Open	1999	Bob Albertson
Des Moines Masters	1999	Al Fredrickson
Indiana Invitational	1999	Chip Masterson

**Winner Dates of Birth**

<u>Winner</u>	<u>Date of Birth</u>
Chip Masterson	14 March 1977
Al Fredrickson	21 July 1975
Bob Albertson	28 September 1968



# Ejercicios!