

MP02-RA 6

Normalització BBDD



Index

1 Introducció

1.1 Disseny incorrecte

2 Teoria de la normalització

3 Formes normals

3.1 Primera forma normal

3.2 Segona forma normal

3.3 Tercera forma normal

3.4 Forma normal de Boyce-Codd

3.5 Quarta forma normal

3.6 Cinquena forma normal

4 Desnormalització

1 Introducció

La **teoria de la normalització** és un mètode que permet assegurar si un disseny relacional (tant si prové de la traducció d'un diagrama Entitat-Relació com si s'ha efectuat directament en el model relacional) és més o menys correcte.

En general, els mals dissenys poden originar les situacions següents:

- ✗ Repetició de la informació
- ✗ Impossibilitat de representar certa informació:
 - ✗ Anomalies en les insercions
 - ✗ Anomalies en les modificacions
 - ✗ Anomalies en els esborraments

Un bon disseny ha d'assolir els següents objectius:

- ➔ Emmagatzemar tota la informació necessària amb el mínim d'informació redundant.
- ➔ Mantenir el mínim de lligams entre les relacions de la base de dades per tal de facilitar-ne la utilització.
- ➔ Millorar la consultabilitat de les dades emmagatzemades.
- ➔ Minimitzar els problemes d'actualització (altes, baixes i modificacions) que poden sorgir en haver d'actualitzar simultàniament dades de diferents relacions.

1.1 Disseny incorrecte

DniProf	NomProfessor	DniAlum	NomAlumne	Edat	Credit	Nota
33.333.333	Joan Finestra	77.777.777	Anna Taula	20	ADBD	4.5
33.333.333	Joan Finestra	88.888.888	Miquel Cadira	19	ADBD	5.7
33.333.333	Joan Finestra	77.777.777	Anna Taula	20	SGBD	6
33.333.333	Joan Finestra	88.888.888	Miquel Cadira	19	SGBD	7
44.444.444	Maria Porta	77.777.777	Anna Taula	20	MET	6
44.444.444	Maria Porta	88.888.888	Miquel Cadira	19	MET	5
44.444.444	Maria Porta	77.777.777	Anna Taula	20	LLC	4
44.444.444	Maria Porta	88.888.888	Miquel Cadira	19	LLC	3

1.1 Disseny incorrecte

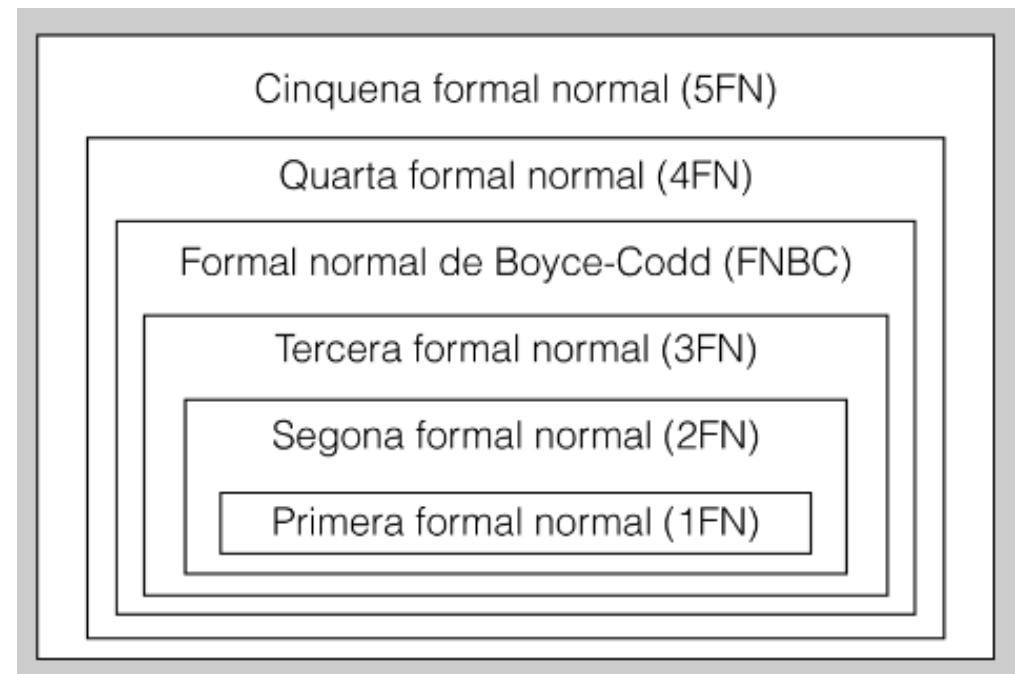
DniProf	NomProfessor	DniAlum	NomAlumne	Edat	Credit	Nota
33-333-333	Joan Finestra	77.777.777	Anna Taula	20	ADBD	4.5
33-333-333	Joan Finestra	88.888.888	Miquel Cadira	19	ADBD	5.7
33-333-333	Joan Finestra	77.777.777	Anna Taula	20	SGBD	6
33-333-333	Joan Finestra	88.888.888	Miquel Cadira	19	SGBD	7
44-444-444	Maria Porta	77.777.777	Anna Taula	20	MET	6
44-444-444	Maria Porta	88.888.888	Miquel Cadira	19	MET	5
44-444-444	Maria Porta	77.777.777	Anna Taula	20	LLC	4
44-444-444	Maria Porta	88.888.888	Miquel Cadira	19	LLC	3

- Hi ha informació repetida, fet que pot provocar inconsistències.
- En cas d'haver de modificar qualsevol dels valors dels camps que formen la clau primària (DniProf, NomProfessor, DniAlum, NomAlumne, Edat, Credit), el canvi s'ha d'efectuar en totes les files en què apareix aquest valor.
- No hi pot haver valors nuls en les columnes que formen la clau primària. Així, si no coneixem Dni d'un alumne, tenim un greu problema.
- En cas d'arribar a la conclusió que necessitem emmagatzemar més informació dels professors o dels alumnes, caldrà afegir més columnes i repetir la informació per a cada fila en què aparegui el professor o alumne.

2 Teoria de la normalització

El mètode que proposa la teoria de la normalització per determinar si un disseny relacional és correcte consisteix a avaluar el disseny de totes les taules per tal de veure en quin grau de normalitat es troba cadascuna i, així, poder decidir si el disseny ja és correcte o si cal refinar-lo.

L'objectiu d'un disseny ha de ser aconseguir un esquema relacional en què totes les taules tingueixin el grau màxim de normalitat, és a dir, en què totes es trobessin en la cinquena forma normal (5FN).



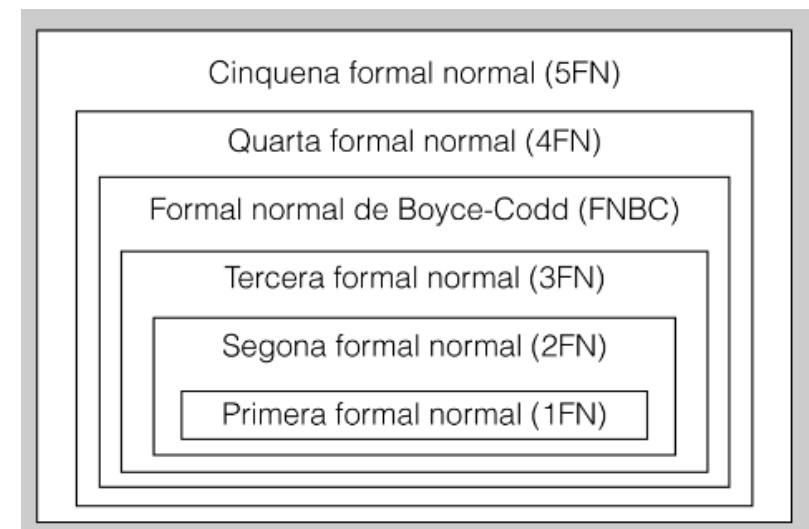
Hi ha diferents nivells(graus) de formes normals, a mesura que augmenta el nivell de la forma normal, la taula ha de complir un conjunt de condicions més restrictiu i, per tant, continua verificant les condicions de les formes normals de nivell inferior.

2 Teoria de la normalització

El **procés de normalització** per aconseguir que una taula que es troba en una forma normal X passi a estar en una forma normal Y superior a X consisteix sempre en la descomposició o subdivisió de la taula original en dues o més taules que verifiquin el nivell de forma normal Y.

Per tant, el procés de normalització augmenta el nombre de taules presents a la BD. Amb això, segur que s'aconsegueix una disminució de redundàncies i una disminució de les anomalies en els problemes d'actualització de la informació, però, en canvi, **es penalitzen les consultes**, ja que per resoldre-les caldrà anar a cercar la informació en moltes taules relacionades entre elles.

Així, doncs, cal trobar un equilibri, i de vegades pot ser convenient renunciar al nivell màxim de normalització (5FN) i, per tant, permetre una certa redundància de dades amb la finalitat d'alleugerir els costos de les consultes. En aquestes situacions, es parla d'un **procés de desnormalització**.



3.1 Primera forma normal

Una taula està en **primera forma normal (1FN)** si cap atribut pot contenir valors no atòmics (indivisible).

TAULA 1

Num	DataComanda	Article	Descripcio	Qtat	Preu	DataPrevista	Moneda
22.523	25-05-2000	PC3-500 PRO-15	PC Pentium III a 500 , Protector Pantalla 15"	5	150 8	1-06-2000	EUR
22.524	27-05-2000	PC3-500 PRO-15	PC Pentium III a 500 , Protector Pantalla 15"	15	145 50	5-06-2000	USD
22.525	27-05-2000	INK430	Cartutx de tinta 430	20	25	31-5-2000	EUR

TAULA 2 (1FN)

Num	DataComanda	Article	Descripcio	Qtat	Preu	DataPrevista	Moneda
22.523	25-05-2000	PC3-500	PC Pentium III a 500	5	150	1-06-2000	EUR
22.523	25-05-2000	PRO-15	Protector pantalla 15"	5	8	1-06-2000	EUR
22.524	27-05-2000	PC3-500	PC Pentium III a 500	15	145	5-06-2000	USD
22.524	27-05-2000	PRO-15	Protector pantalla 15"	15	50	5-06-2000	USD
22.525	27-05-2000	INK430	Cartutx de tinta 430	20	25	31-5-2000	EUR

A la **taula 1** tenim comandes, cada fila representa una comanda i hi ha columnes que contenen més d'un valor per la mateixa fila.

El procés que s'ha de seguir per assolir una 1FN és afegir tantes files com sigui necessari per a cadascun dels diferents valors del camp o camps que tinguin valors no atòmics.

A la **taula 2** han aparegut 2 files més i totes les columnes de totes les files només tenen 1 valor.

En la majoria de dissenys si se segueixen els criteris de definició del model relacional aquest problema no es produirà mai.

3.2 Segona forma normal (I)

Una taula està en **segona forma normal (2FN)** si està en 1FN i tot atribut que no pertany a la clau té dependència funcional total de la clau.

TAULA (1FN)

Num	DataComanda	Article	Descripcio	Qtat	Preu	DataPrevista	Moneda
22.523	25-05-2000	PC3-500	PC Pentium III a 500	5	150	1-06-2000	EUR
22.523	25-05-2000	PRO-15	Protector pantalla 15"	5	8	1-06-2000	EUR
22.524	27-05-2000	PC3-500	PC Pentium III a 500	15	145	5-06-2000	USD
22.524	27-05-2000	PRO-15	Protector pantalla 15"	15	50	5-06-2000	USD
22.525	27-05-2000	INK430	Cartutx de tinta 430	20	25	31-5-2000	EUR

A la taula de comandes, la clau la formen **Num** i **Article**, la resta d'atributs de la taula han d'estar relacionades amb la clau, tota la clau, per complir la 2ona forma normal. Això no passa per exemple amb **Descripció** que només depèn d'una part de la clau, d'**Article**, i no té res a veure amb **Num** de comanda.

El procés que s'ha de seguir per assolir una 2FN és dividir la taula (conservant la informació i les dependències) en tantes taules com sigui necessari de manera que cada taula verifiqui que els seus atributs no-clau tenen dependència funcional total de la clau.

3.2 Segona forma normal (II)

Una taula està en **segona forma normal (2FN)** si està en 1FN i tot atribut que no pertany a la clau té dependència funcional total de la clau.

TAULES EN ·2FN

COMANDA			
Num	DataComanda	DataPrevista	Moneda
22.523	25-05-2000	1-06-2000	EUR
22.524	27-05-2000	5-06-2000	USD
22.525	27-05-2000	31-5-2000	EUR

ARTICLE	
Article	Descripcio
PC3-500	PC Pentium III a 500
PRO-15	Protector pantalla 15"
INK430	Cartutx de tinta 430

DETALL			
Num	Article	Qtat	Preu
22.523	PC3-500	5	150
22.523	PRO-15	5	8
22.524	PC3-500	15	145
22.524	PRO-15	15	50
22.525	INK430	20	25

Per aconseguir unes taules que compleixen **2FN** hem dividit la taula original en 3 taules.

- A la taula **COMANDA** hem posat els atributs que defineixen una comanda, depenents de **Num**.
- A la taula **ARTICLE** hem posat els atributs corresponents a un article
- A la taula **DETALL** hem posat els atributs que depenen de **Num** i **Article**.

En la majoria de dissenys si se segueixen els criteris de definició del model ER aquest problema no es produeix.

3.3 Tercera forma normal (I)

Una taula està en **tercera forma normal (3FN)** si està en 2FN i cap atribut que no pertany a la clau primària, ni forma part d'una clau candidata, depèn **transitivament** (funcionalment) de la clau primària.

COMANDA					
Num	Data	DataPrevista	NomProv	PaisProv	Moneda
22.523	25-05-2000	25-05-2000	ARKANSAS	XINA	EUR
22.524	27-05-2000	5-06-2000	MELISSA	ITÀLIA	USD
22.525	27-05-2000	31-5-2000	ARKANSAS	XINA	EUR

TAULA EN 2FN

COMANDA				
Num	DataComanda	DataPrevista	CodProv	Moneda
22.523	25-05-2000	1-06-2000	ARK	EUR
22.524	27-05-2000	5-06-2000	MEL	USD
22.525	27-05-2000	31-5-2000	ARK	EUR

TAULES EN 3FN

PROVEIDOR		
CodProv	NomProv	PaisProv
ARK	ARKANSAS	XINA
MEL	MELISSA	ITÀLIA

A la taula 1 l'atribut **PaisProv** no depèn transitivament de **Num** (PK), sinò de **NomProv**. Per tant, aquesta taula no està en 3FN.

El procés que s'ha de seguir per assolir una 3FN és dividir la taula (conservant la informació i les dependències) en noves taules més simples, de manera que en cada taula es compleixi que els seus atributs no-clau depèn transitivament de la clau.

Al fer la conversió a 3FN hem aprofitat per definir una nova clau primària per PROVEIDOR, en comptes del nom que la clau sigui un codi. L'experiència aconsella definir un codi que ens permeti identificar-los de manera més clara que la que proporciona el seu nom.

3.3 Tercera forma normal (II)

Què es una DEPENDÈNCIA TRANSITIVA?

A les bases de dades relacionals, la dependència transitiva es produeix quan un atribut depèn indirectament d'un altre mitjançant un atribut intermedi.

Agafem com exemple la següent taula, l'atribut intermedi és instructor:

Curso	Instructor	Instructor_titulo
Manipulació de datos en SQL	Paco Vázquez	Científico de datos
Fundamentos de SQL	Paco Vázquez	Científico de datos
Introducció a SQL	Milagros Ortega	Entrenador de datos

D'aquesta taula les dependències són:

- *Curso* → *Instructor*: Si coneixem el curs, coneixem l'instructor.
- *Instructor NO* → *Curso*: De vegades, els instructors ensenyan més d'un curs.
- *Instructor* → *Instructor_titulo*: Cada instructor té un títol.

Llavors, *Curso* → *Instructor_titulo* és una dependència transitiva.

Per més informació teniu el següent link: [transitive-dependency](#)

3.4 FN de Boyce-Codd

EXEMPLE:

Considerem les naus d'emmagatzematge que hi ha en un gran mercat dedicades a guardar les mercaderies dels venedors del mercat. Imaginem que:

- cada nau guarda només mercaderia d'un tipus concret (carn fresca, peix fresc, congelats, vegetals, basar...)
- cada venedor pot dipositar mercaderia en diferents naus segons el tipus de mercaderia de cada nau.
- tota la mercaderia d'unes mateixes característiques d'un venedor es troba concentrada en una ubicació dins la mateixa nau.

DIPOSIT

Venedor	TipusMaterial	Nau	Ubicacio
JOSEP	Peix fresc	15	S4-C3-U _{1:10}
MARIA	Peix fresc	25	S3-C5-U _{5:22}
RAMON	Congelats	17	S2-C4-U _{1:25}
ANNA	Vegetals	20	S1-C6-U _{7:10}
ANNA	Peix fresc	25	S2-C5-U _{12:15}
MARIA	Basar	10	S3-C4-U _{20:25}

La taula ens mostra la situació. Fixem-nos que es troba en 3FN:

- En efecte, la **Nau** i la **Ubicacio** dins la nau depenen del Venedor i del TipusMaterial, ja que hi pot haver diverses naus dedicades a un tipus de material, però tot el material similar d'un venedor es troba en una nau determinada.
- Al mateix temps, hi pot haver diverses naus amb material d'un venedor a causa de la different tipologia del material.

3.4 FN de Boyce-Codd

Aquesta taula, tot i estar en 3FN, presenta anomalies:

- Si en un moment donat una nau no té material de cap venedor, es perd la informació referent al tipus de mercaderia que correspon a la nau.
- Si canvia la descripció del tipus de mercaderia assignada a una nau, cal modificar tantes files com venedors amb dipòsits d'aquell tipus de mercaderia hi hagi a la nau.

Una taula està en la forma normal de Boyce-Codd (**FNBC**) si està en **2FN** i tots els seus **determinants** són claus candidates.

*El procés que s'ha de seguir per assolir una **FNBC** és apartar de la taula els atributs que depenen de determinants que no són claus candidates, i crear noves taules que recullen aquests atributs*

DIPOSIT				
Venedor	TipusMaterial	Nau	Ubicacio	
JOSEP	Peix fresc	15	S4-C3-U1:10	
MARIA	Peix fresc	25	S3-C5-U5:22	
RAMON	Congelats	17	S2-C4-U1:25	
ANNA	Vegetals	20	S1-C6-U7:10	
ANNA	Peix fresc	25	S2-C5-U12:15	
MARIA	Basar	10	S3-C4-U20:25	

La taula DIPOSIT no es troba en FNBC, ja que l'atribut **Nau** determina el **TipusMaterial** però **Nau** no és clau candidata de la taula DIPOSIT.

DIPOSIT			NAU_TIPUS	
Venedor	Nau	Ubicacio	Nau	TipusMaterial
JOSEP	15	S4-C3-U1:10	10	Basar
MARIA	25	S3-C5-U5:22	15	Peix fresc
RAMON	17	S2-C4-U1:25	17	Congelats
ANNA	20	S1-C6-U7:10	20	Vegetals
ANNA	25	S2-C5-U12:15	25	Peix fresc
MARIA	10	S3-C4-U20:25		

3.5 Quarta forma normal (I)

Una taula està a **4FN** si està en 3FN o FNBC i no posseeix dependències multivalor no trivials. Una taula amb una dependència multivalor és una on l'existència de dues o més relacions independents molts a molts causa redundància

<u>Restaurante</u>	<u>Variedad de Pizza</u>	<u>Área de envío</u>
Vincenzo's Pizza	Corteza gruesa	Springfield
Vincenzo's Pizza	Corteza gruesa	Shelbyville
Vincenzo's Pizza	Corteza fina	Springfield
Vincenzo's Pizza	Corteza fina	Shelbyville
Elite Pizza	Corteza fina	Capital City
Elite Pizza	Corteza rellena	Capital City
A1 Pizza	Corteza gruesa	Springfield
A1 Pizza	Corteza gruesa	Shelbyville
A1 Pizza	Corteza gruesa	Capital City
A1 Pizza	Corteza rellena	Springfield
A1 Pizza	Corteza rellena	Shelbyville
A1 Pizza	Corteza rellena	Capital City

EXEMPLE:

A la taula de l'esquerra cada fila indica que un restaurant donat pot lliurar una varietat donada de pizza a una àrea donada.

A la **taula tots els atributs formen part de la clau única**, no hi ha cap atribut no-clau, no viola cap forma normal fins a l'FNBC.

Però pel fet que les varietats de pizza que un restaurant ofereix són independents de les àrees a les quals el restaurant envia, hi ha redundància a la taula:

- per exemple, ens diuen tres vegades que «A1 Pizza» ofereix la «corteza rellena», i si A1 Pizza comença a produir pizzas de «corteza con queso» llavors necessitarem afegir múltiples registres, un per a cadascuna de les àrees d'enviament d'A1 Pizza.

En termes formals, això es descriu com que «Varietat de pizza» està tenint una dependència multivalor a «Restaurant».

3.5 Quarta forma normal (II)

Una taula està a **4FN** si està en 3FN o FNBC i no posseeix dependències multivalor no trivials. Una taula amb una dependència multivalor és una on l'existència de dues o més relacions independents molts a molts causa redundància

TAULA EN 3FN

Restaurante	Variedad de Pizza	Área de envío
Vincenzo's Pizza	Corteza gruesa	Springfield
Vincenzo's Pizza	Corteza gruesa	Shelbyville
Vincenzo's Pizza	Corteza fina	Springfield
Vincenzo's Pizza	Corteza fina	Shelbyville
Elite Pizza	Corteza fina	Capital City
Elite Pizza	Corteza rellena	Capital City
A1 Pizza	Corteza gruesa	Springfield
A1 Pizza	Corteza gruesa	Shelbyville
A1 Pizza	Corteza gruesa	Capital City
A1 Pizza	Corteza rellena	Springfield
A1 Pizza	Corteza rellena	Shelbyville
A1 Pizza	Corteza rellena	Capital City

Per satisfer la **4FN**, hem de posar la informació de les varietats de pizza ofertes en una taula diferent a la de la informació sobre àrees d'enviament:

TAULES EN 4·FN

Restaurante	Variedad de pizza
Vincenzo's Pizza	Corteza gruesa
Vincenzo's Pizza	Corteza fina
Elite Pizza	Corteza fina
Elite Pizza	Corteza rellena
A1 Pizza	Corteza gruesa
A1 Pizza	Corteza gruesa
A1 Pizza	Corteza gruesa
A1 Pizza	Corteza rellena
A1 Pizza	Corteza rellena
A1 Pizza	Corteza rellena

Restaurante	Área de envío
Vincenzo's Pizza	Springfield
Vincenzo's Pizza	Shelbyville
Elite Pizza	Capital City
A1 Pizza	Springfield
A1 Pizza	Shelbyville
A1 Pizza	Capital City

3.6 Cinquena forma normal (I)

Una taula es diu que està en **5FN** si i només si està en 4FN i cada dependència d'unió (join) que hi ha és implicada per les claus candidates.

Psiquiatra	Asegurador	Condición
Dr. James	Healthco	Ansiedad
Dr. James	Healthco	Depresión
Dr. Kendrick	FriendlyCare	OCD
Dr. Kendrick	FriendlyCare	Ansiedad
Dr. Kendrick	Healthco	Depresión
Dr. Lowenstein	FriendlyCare	Esquizofrenia
Dr. Lowenstein	Healthco	Ansiedad
Dr. Lowenstein	Healthco	Demencia
Dr. Lowenstein	Victorian Life	Trastorno de conversión

EXEMPLE:

El psiquiatre pot oferir tractament als pacients que pateixen la condició donada i que són assegurats per l'assegurador donat.

En absència de qualsevol regla, donada per l'escenari real que es vol representar, que restringeixi les combinacions vàlides possibles de psiquiatre, assegurador, i condició, la taula de tres atributs Psiquiatre-per-Assegurador-per-Condició és necessària per modelar la situació correctament.

L'única clau candidata que permet identificar de manera unívoca (sense duplicats) cadascuna de les files de la taula és la unió de les tres columnes Psiquiatre+Assegurador+Condició i, en conseqüència, aquesta n'és la clau primària composta (PK).

La taula serà vàlida i complirà amb la 5FN si en l'escenari real que es vol representar hi ha limitacions a la relació contractual d'un doctor amb un assegurador, com per exemple:

- *L'assegurador FriendlyCare, podria no permetre al Dr. Lowenstein (malgrat tenir-ne la capacitat) donar tractament per a l'Ansietat sota la seva cobertura, però sí li permetria donar tractament per a l'Esquizofrènia. FriendlyCare ja ofereix el servei de tractament per a l'Ansietat amb un altre doctor.*

3.6 Cinquena forma normal (II)

Clau primaria es la combinació de estàs 3 en la primera taula

<u>Psiquiatra</u>	<u>Asegurador</u>	<u>Condición</u>
Dr. James	Healthco	Ansiedad
Dr. James	Healthco	Depresión
Dr. Kendrick	FriendlyCare	OCD
Dr. Kendrick	FriendlyCare	Ansiedad
Dr. Kendrick	Healthco	Depresión
Dr. Lowenstein	FriendlyCare	Esquizofrenia
Dr. Lowenstein	Healthco	Ansiedad
Dr. Lowenstein	Healthco	Demencia
Dr. Lowenstein	Victorian Life	Trastorno de conversión

Model 1

EXEMPLE (cont):

Suposem que l'escenari és un altre i que respon a la regla lògica següent:

- un psiquiatre pot tractar la condició C.
- una asseguradora cobreix la condició C
- un psiquiatre és autoritzat a oferir el tractament reemborsable als pacients assegurats per l'assegurador A
- aleshores, en cas que l'assegurador A cobreixi la condició C - ha de ser cert que el psiquiatre pot oferir el tractament als pacients que pateixen la condició C i estan assegurats per l'assegurador A.

Si es compleixen la regla anterior, és possible dividir la taula en tres parts per emmagatzemar tots els parells de combinacions possibles:

Per a les tres noves taules la totalitat de les columnes definiran la clau primària composta (PK) ja que és l'única clau candidata que existeix per a cada taula.

Psiquiatra-para-Condición	
<u>Psiquiatra</u>	<u>Condición</u>
Dr. James	Ansiedad
Dr. James	Depresión
Dr. Kendrick	OCD
Dr. Kendrick	Ansiedad
Dr. Kendrick	Depresión
Dr. Lowenstein	Esquizofrenia
Dr. Lowenstein	Ansiedad
Dr. Lowenstein	Demencia
Dr. Lowenstein	Trastorno de conversión

Psiquiatra-para-Asegurador		Asegurador-para-Condición	
<u>Psiquiatra</u>	<u>Asegurador</u>	<u>Asegurador</u>	<u>Condición</u>
Dr. James	Healthco	Healthco	Ansiedad
Dr. Kendrick	FriendlyCare	Healthco	Depresión
Dr. Kendrick	Healthco	Healthco	Demencia
Dr. Lowenstein	FriendlyCare	FriendlyCare	Ansiedad
Dr. Lowenstein	Healthco	FriendlyCare	Esquizofrenia
Dr. Lowenstein	Victorian Life	Victorian Life	Trastorno de conversión

Model 2

Aquesta disposició ajuda a treure redundància (dades duplicades). També facilita les actualitzacions de dades:

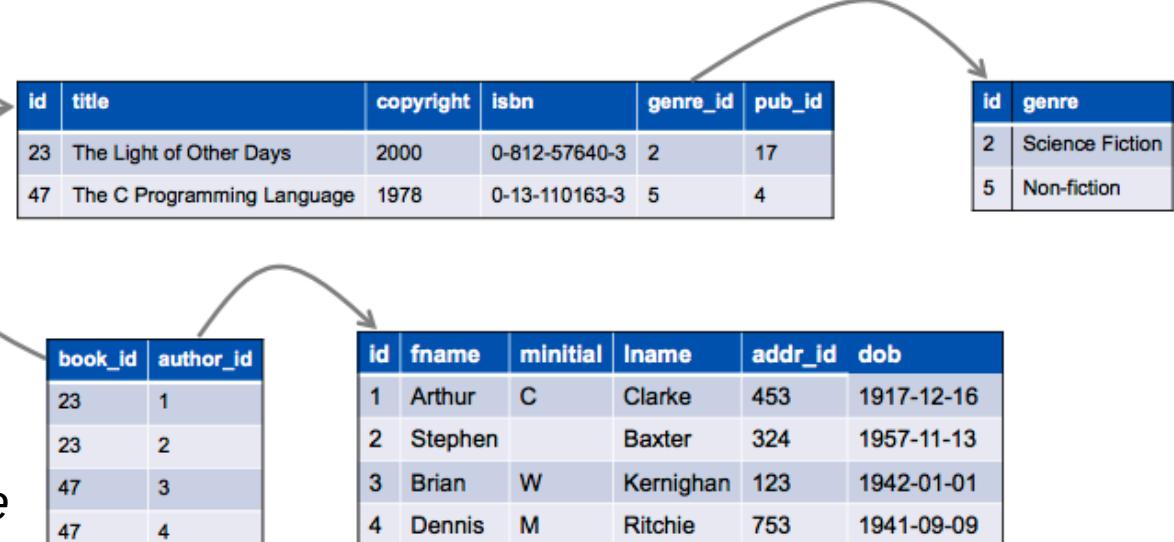
- Supposeu que el Dr. James es converteix en un proveïdor de tractaments per a Friendly Care. Amb el model 1 hauríem d'afegir dues noves entrades ja que el Dr. James pot tractar dues condicions cobertes per Friendly Care: ansietat i depressió. Amb el model 2 necessitem afegir una sola entrada (a la taula Psiquiatre-per-Assegurador).

4 Desnormalització(I)

La **desnormalització** és pot definir com la introducció de redundàncies de forma controlada en una bases de dades, per tal de fer més eficients alguns processos que, altrament, farien que globalment el rendiment del sistema resultés poc òptim.

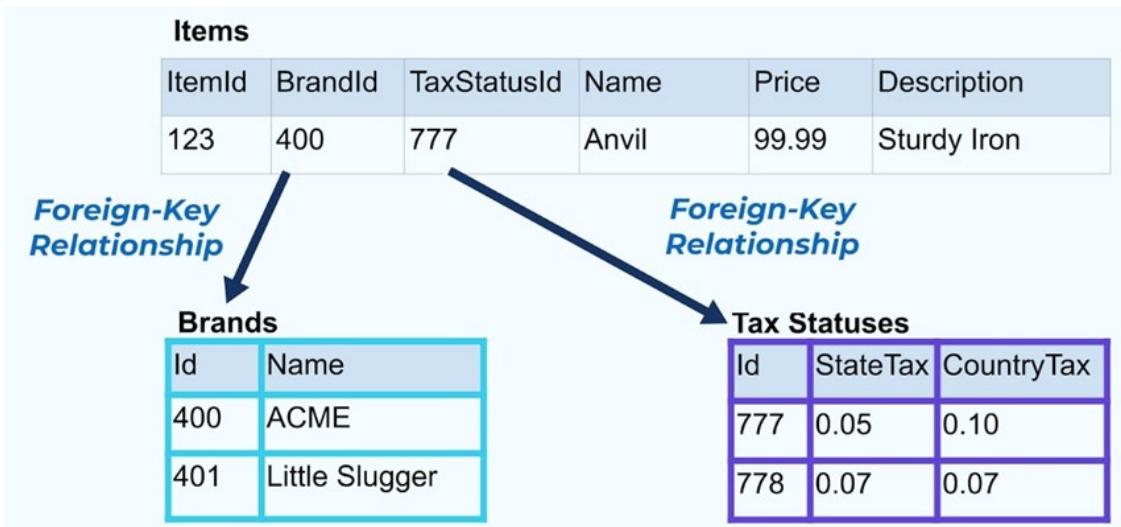
Denormalization is the process of straying away from normalization rules to speed up complex queries. The underlying principle is to strategically duplicate and group data in ways that optimize read operations.

While it may sound counterintuitive to introduce redundancy intentionally, the benefits in terms of speed are impressive. A report by Oracle showed that denormalized databases could improve query response times by up to 100%.



This does not mean denormalization is a one-size-fits-all solution. It should be used judiciously only in scenarios where read performance is more critical than data storage efficiency.

4 Desnormalització(II)



Tant la taula de *brands* com la de *tax status* estan relacionades amb la taula de *items*. Tot i que només tenim un *item* a la taula d'articles, podeu imaginar que hi pot haver milers o milions d'articles, i que molts d'aquests articles tindran les mateixes propietats de *brand* i *tax status*.

En una base de dades desnormalitzada, aquesta estructura es podria modificar i convertir-se en la següent:

Denormalized Items						
ItemId	BrandName	StateTax	CountryTax	Name	Price	Description
123	ACME	0.05	0.10	Anvil	99.99	Sturdy Iron

Diguem que l'empresa ACME fa un canvi de marca i crea un nou nom de marca:

- canvia d'ACME a Rotunda.

Quines són les conseqüències del canvi en aquesta BD desnormalitzada?

4 Desnormalització(III)

EXEMPLES DE DESNORMALITZACIÓ

- En BD on hi ha consultes intensives sobre dades d'una taula que tenen referenciades altres taules on s'hi emmagatzemen descripcions, en algun cas pot ser útil mantenir descripció i codi en la mateixa taula, per tal de fer més ràpida la consulta mencionada. Per exemple, una taula d'adreces d'empleats que té el codi de la província i que fa referència a una altra taula que té la descripció de la província. Cada cop que calgui obtenir la província dels empleats caldrà processar l'operació de join de les dues taules. En canvi, si es disposa del nom de la província en la taula d'adreces, s'evitarà aquesta operació que resulta costosa.
- Afegir camps que són calculables, com ara el total d'una factura, pot reduir, també, el temps de consulta de les dades d'una factura.

4 Desnormalització (IV)

Tot i que per a un sistema donat, es podrien preveure certes modificacions sobre la BD per tal de millorar el rendiment abans de la posada en funcionament, el més típic és detectar aquestes necessitats a posteriori.

Així, doncs, aquests tipus de modificacions sobre la base de dades es duen a terme, habitualment, després d'haver observat certes ineficiències en algunes operacions habituals sobre la base de dades.

La introducció de redundàncies que desnormalitzin la BD habitualment va lligat a la implementació de certs mecanismes que intenten solucionar els possibles problemes que es podrien generar d'aquestes accions. La forma més típica de controlar els canvis per a evitar inconsistències en aquests casos és la creació de triggers.

Denormalization vs. Normalization

	Denormalization	Normalization
PROS	<ul style="list-style-type: none">+ Faster data reads+ Simpler queries advantageous to developers+ Requires less compute on read operations+ Makes data available quickly	<ul style="list-style-type: none">+ Faster writes+ No redundant data+ Less database complexity+ Data always consistent
CONS	<ul style="list-style-type: none">- Slower writes- More database complexity- Potential for data inconsistency- Requires more storage, RAM	<ul style="list-style-type: none">- Slower reads- Heavy querying can overwhelm, crash hardware- Table joins required since data isn't duplicated- Indexing not as efficient due to table joins

normalització vs desnormalització

Es pot programar un trigger, per exemple, perquè en inserir, modificar o eliminar una línia de factura, es recalculi i actualizi l'import del total de la factura, per així tenir aquest camp calculat sempre consistent.

Precs i pregunes

