

الفصل السادس: التنظيم (Normalization) (1)



Database Architecture And Design $_$ CH6 $^{\circ}$

العنوان	الصفحة
1. مقدمة	3
2. إشكاليات التصميم السيئ لقاعدة البيانات	3
3. مستويات التنظيم	6
1.3 الشكل النظامي الأول (1NF)	7
2.3 الشكل النظامي الثاني (2NF)	10
3.5 الشكل النظامي الثالث (3NF)	13
4. تدریبات	15

ملخص

يركز هذا الفصل على تعريف إشكاليات التصميم السيئ لقاعدة البيانات، وطرق تنظيم القاعدة.

الأهداف التعليمية

يتضمن هذا الفصل:

- تعريف إشكاليات التصميم السيئ لقاعدة البيانات
- اشكالية الادخال (Insertion anomaly).
- المكالية الحذف (Deletion anomaly).
- الشكالية التعديل (Update anomaly).
 - الشكل النظامي الأول (1NF)
 - الشكل النظامي الثاني (2NF)
 - الشكل النظامي الثالث (3NF)

1. مقدمة

تعرفنا سابقاً على مستويات التجريد الثلاثة التي يمر بها بناء قاعدة البيانات، وهي المستوى المفهومي والمستوى المنطقي والمستوى الفيزيائي، ورأينا أن المستوى المفهومي ينتهي ببناء مخطط ERD ويتضمن ما يلي:

- عزل الكيانات.
- تحديد الواصفات ومجالات تعريفها
 - تحديد العلاقات بين الكيانات
- تحديد المفاتيح المرشحة والمفاتيح الأساسية للكيانات
 - إنشاء مخطط ERD
- تدقيق مخطط ERD من قبل مصمم قاعدة البيانات بالتعاون مع المستثمرين

يبدأ المستوى المنطقي في بناء قاعدة المعطيات بعملية التنظيم، ويقصد بالتنظيم (Normalization) تصميم جداول قاعدة البيانات بحيث نتحكم بتكرار المعطيات ونتجنب حالات الشذوذ التي يمكن أن تنتج عن عمليات الإضافة والحذف والتعديل على المعطيات.

سنقوم بداية بدراسة الإشكاليات التي يعاني منها التصميم السيئ لقاعدة البيانات، ومن ثم سندرس مستويات التنظيم وخصائص كل منها.

2. إشكاليات التصميم السيئ لقاعدة البيانات

بفرض أن شركة ما تضم مجموعة من مسؤولي المبيعات يعملون في عدة مواقع، وتريد الإدارة تخزين البيانات المتعلقة بموظفيها وبيانات مستودعاتها في قاعدة معطيات. المحاولة الأولى لبناء القاعدة نتج عنها الجدول التالى:

ld	Name	Address	Title	Store	Store_Address	Store_Phone
Sr1	Jane	E1	Sales Rep	A1	S1	011-1234567
Sr2	Fred	E2	Sales Rep	A1	S1	011-1234567
Sr3	Ed	E3	Manager	A1	S1	011-1234567
Sr4	Ann	E4	Sales Rep	A2	S2	011-7891234
Sr5	Jone	E5	Sales Rep	A2	S2	011-7891234
Sr6	Smith	E6	Manager	A2	S2	011-7891234

 $^{-}$ يتضمن هذا الجدول 1 بيانات الموظفين وبيانات المستودعات، ومن الواضح وجود تكرار في بيانات المستودعات. إذا أردنا إدخال بيانات موظف جديد فيجب إدخال بيانات موقعه (المستودع) أيضاً، وللحفاظ على عدم تناقض البيانات في القاعدة يجب إدخال بيانات المستودع بدقة وبشكل مطابق للقيم المدخلة سابقاً (لأحد الموظفين السابقين في نفس المستودع)، هنا تظهر مشكلة تكرار البيانات بالإضافة طبعاً لحجم التخزين المهدور.

1. بفرض أننا نريد إدخال بيانات مستودع جديد قبل توظيف أحد فيه، هذا يتطلب إدخال قيم Null في بيانات الموظف ومن ضمنها الحقل المفتاح (ld) مما يخرق شرط تكامل البيانات. هذا ما يدعى بإشكالية الإدخال (Insertion anomaly).

ld	Name	Address	Title	Store	Store_Address	Store_Phone
Sr1	Jane	E1	Sales Rep	A 1	s ₁	011-1234567
Sr2	Fred	E 2	Sales Rep	A 1	S ₁	011-1234567
Sr3	Ed	E3	Manager	A 1	S 1	011-1234567
Sr4	Ann	E4	Sales Rep	A 2	S 2	011-7891234
Sr5	Jone	E 5	Sales Rep	A 2	S 2	011-7891234
Sr6	Smith	E6	Manager	A 2	S 2	011-7891234
Null	Null	Null	Null	A 3	S 3	011-9123456

2. بفرض أننا حذفنا تسجيلة الموظف الأخير في أحد المستودعات، سيؤدي ذلك إلى فقدان بيانات مستودع موجود في الشركة. هذا ما يدعى بإشكالية الحذف (Deletion anomaly).

ld	Name	Address	Title	Store	Store_Address	Store_Phone
Sr1	Jane	E1	Sales Rep	A1	S ₁	011-1234567
Sr2	Fred	E 2	Sales Rep	A 1	S ₁	011-1234567
Sr3	Ed	E3	Manager	A1	S1	011-1234567



⚠ Store A2 data is missing

3. إذا أردنا تغيير رقم هاتف أحد المستودعات، عندها يجب تغيير رقم الهاتف في كل تسجيلات موظفي ذلك المستودع، فإذا تم تغيير جزء من هذه التسجيلات دون جزء آخر سيؤدي ذلك إلى بيانات متناقضة في قاعدة البيانات، وهذا ما يدعى بإشكالية التعديل (Update anomaly).

ld	Name	Address	Title	Store	Store_Address	Store_Phone
Sr1	Jane	E1	Sales Rep	A 1	s ₁	011-1234567
Sr2	Fred	E 2	Sales Rep	A 1	S ₁	011-1234567
Sr3	Ed	E3	Manager	A 1	S 1	011-1234567
Sr4	Ann	E4	Sales Rep	A 2	S 2	011-3457890
Sr5	Jone	E 5	Sales Rep	A 2	S 2	011-7891234
Sr6	Smith	E6	Manager	A 2	S 2	011-7891234



حل المشاكل السابقة يتم من خلال تجزيء الجدول السابق إلى جدولين، أحدهما يحمل بيانات الموظفين ويستورد مفتاح الجدول الثاني الذي يحمل بيانات المستودعات:

ld	Name	Address	Title	Store	Store_Address	Store_Phone
Sr1	Jane	E1	Sales Rep	A1	S1	011-1234567
Sr2	Fred	E2	Sales Rep	A1	S1	011-1234567
Sr3	Ed	E3	Manager	A1	S1	011-1234567
Sr4	Ann	E4	Sales Rep	A2	S2	011-7891234
Sr5	Jone	E5	Sales Rep	A2	S2	011-7891234
Sr6	Smith	E6	Manager	A2	S2	011-7891234

ld	Name	Address	Title	Store
Sr1	Jane E1 Sales Rep		Sales Rep	A1
Sr2	Fred E2		Sales Rep	A1
Sr3	Ed	E3	Manager	A1
Sr4	Ann	E4	Sales Rep	A2
Sr5	Jone	E5	Sales Rep	A2
Sr6	Smith E6		Manager	A2

Store	Store_Address	Store_phone
A1	S1	011-1234567
A2	S2	011-7891234

بذلك فإنه:

- 1. عند إدخال بيانات موظف جديد، يتم تسجيل مفتاح المستودع الذي يعمل فيه، أما بقية بيانات المستودع فهي مخزنة في جدول مستقل وبالتالي لن تظهر بيانات متناقضة في قاعدة البيانات بشأن المستودعات، وإذا أردنا إدخال بيانات مستودع جديد فيمكن إدخالها بصرف النظر عمن يعمل في المستودع، مما يجنبنا إشكالية الإدخال.
 - 2. عند حذف الموظف الأخير في مستودع، تبقى بيانات المستودع في جدول مستقل.
 - 3. عند تعديل رقم هاتف المستودع، يتم التعديل على تسجيلة واحدة في جدول المستودعات.

الإجراءات السابقة (تجزيء الجدول) هي عملية تنظيم قاعدة البيانات، والتي سنتعرف عليها بتفصيل أكثر في هذا الفصل والفصل القادم.

3. مستويات التنظيم

كما سبق وعرفنا، التنظيم هو عملية ترتيب وتوزيع جداول قاعدة المعطيات العلائقية، للتقليل من تكرار المعطيات وتقليص حجم التخزين المطلوب ولحل إشكاليات الإدخال والحذف والتعديل. وغالباً ما ينتج عن عملية التنظيم زيادة في عدد جداول القاعدة.

يتم تنظيم قاعدة البيانات من خلال إخضاعها لمجموعة اختبارات والتعديل على البنية لتحقيق مجموعة معايير. مستويات التنظيم الأساسية هي ثلاث (الشكل النظامي الأول 1NF، الشكل النظامي الثالث 3NF)، ويقصد عادةً بتنظيم القاعدة وضعها في الشكل النظامي الثالث.

يضاف إلى المستويات الثلاثة الشكل المقترح من قبل R.Boyce ويطلق عليه تسمية BCNF، والشكلين النظاميين الرابع والخامس.

سنتناول في هذا الفصل الأشكال النظامية الثلاثة الأساسية، وسنتناول المستويات BCNF, 4NF, 5NF في الفصل اللحق.

1.3 الشكل النظامي الأول (1NF)

يقال عن جدول في قاعدة البيانات أنه من الشكل النظامي الأول إذا كان تقاطع كل سطر وعمود فيه، يتضمن قيمة وحيدة غير قابلة للتجزئة.

مثال: يمثل الشكل التالي بيانات شركة تأجير عقارات:

C_id	P_num	C_name	P_address	R_start	R_end	Rent	O_num	Owner
0.1	Pr3	lana	Α	1-1-96	12-1-98	785	Po23	Jones
01	Pr22	Jane	В	2-1-98	3-30-00	1200	Po44	Jan
02	Pr17	Fred	С	2-1-88	1-11-90	1000	Po32	Jill
03	Pr32	Ed	D	6-1-90	3-1-95	950	Po32	Jill
03	Pr22	Lu	В	4-1-00	null	1200	Po44	Jan

رقم الزبون	² C_id
رقم العقار	P_num
اسم الزبون	C_name
عنوان العقار	P_address
تاريخ بدء الإيجار	R_start
تاريخ نهاية الإيجار	R_end
قيمة الإيجار	Rent
رقم مالك العقار	O_num
اسم مالك العقار	Owner

² هو المفتاح الأساسي للجدول

مثال (تتمة):

يمكن ملاحظة أن قيم خلايا الجدول قابلة للتجزئة، وهذا التصميم سيء حيث لا يمكن فيه استخلاص معلومات مالك عقار معين بسهولة. يعاني التصميم أيضاً من مشكلة تكرار المعطيات فكلما تم تأجير عقار يجب أن تدخل بياناته التفصيلية (بيانات العقار) من جديد.

C_id	P_num	C_name	P_address	R_start	R_end	Rent	O_num	Owner
0.1	Pr3		Α	1-1-96	12-1-98	785	Po23	Jones
01 Pr22 Ja	Jane	В	2-1-98	3-30-00	1200	Po44	Jan	
02	Pr17	Fred	С	2-1-88	1-11-90	1000	Po32	Jill
03 Pr32 Pr22	Pr32	Ed	D	6-1-90	3-1-95	950	Po32	Jill
	Pr22	Ed	В	4-1-00	null	1200	Po44	Jan

مثال (تتمة):

لوضع الجدول السابق في الشكل النظامي الأول يجب تجزئة الخلايا الحاوية على معطيات قابلة للتجزئة، يمكن عمل ذلك من خلال إدخال بيانات كل عملية إيجار في سطر جديد، أي يجب أن يصبح المفتاح الأساسي هو نتيجة تركيب رقم الزبون ورقم العقار (c_id, p_num) علماً أن التصميم مبني على أساس تخزين بيانات العقارات المؤجرة حالياً دون حفظ تاريخ حركات الإيجار، بمعنى أن نفس الزبون لا يمكن أن تتواجد له حركتي إيجار لنفس العقار.

C_id	P_num	C_name	P_address	R_start	R_end	Rent	O_num	Owner
01	Pr3	Jane	Α	1-1-96	12-1-98	785	Po23	Jones
	Pr22		В	2-1-98	3-30-00	1200	Po44	Jan
02	Pr17	Fred	O	2-1-88	1-11-90	1000	Po32	Jill
03	Pr32		D	6-1-90	3-1-95	950	Po32	Jill
	Pr22	Ed	В	4-1-00	null	1200	Po44	Jan

C_id	P_num	C_name	P_address	R_start	R_end	Rent	O_num	Owner
01	Pr3	Jane	А	1-1-96	12-1-98	785	Po23	Jones
01	Pr22	Jane	В	2-1-98	3-30-00	1200	Po44	Jan
02	Pr17	Fred	С	2-1-88	1-11-90	1000	Po32	Jill
03	Pr32	Ed	D	6-1-90	3-1-95	950	Po32	Jill
03	Pr22	Ed	В	4-1-00	null	1200	Po44	Jan

تطبيق: طبق معيار 1NF على الجدول التالي:

id	name	children	birth_date
1001	John Doe	Betty, Frank	2-2-88, 4-3-90
1002	Jane Doe	Betty, Frank	2-2-88, 4-3-90
1003	Freda Fish	Henry,	4-4-79, 2-8-84,
		Jane, Jill,	7-9-88, 10-3-90
		Bill	
1004	Bill Bass	Hank, April,	5-4-89, 9-9-94,
		Ellen	7-10-98

الحل:

id f_name		I_name
1001	John	Doe
1002	Jane	Doe
1003	Freda	Fish
1004	Bill	Bass

id	name	birth_date
1001	Betty	2-2-88
1001	Frank	4-3-90
1002	Betty	2-2-88
1002	Frank	4-3-90
1003	Henry	4-4-79
1003	Jane	2-8-84
1003	Jill	7-9-88
1003	Bill	10-3-90
1004	Hank	5-4-89
1004	April	9-9-94
1004	Ellen	7-10-98

وضع الجدول في الشكل النظامي الأول لا يحل مشاكل التصميم السيئ إنما هو خطوة في طريق حلها، فبالرغم من وضع جدول إيجار العقارات في الشكل النظامي الأول فإنه لا يزال يعاني من مشكلة تكرار البيانات، وإشكاليات الإضافة والتعديل والحذف.

2.3 الشكل النظامي الثاني (2NF)

يقال عن جدول أنه من الشكل النظامي الثاني إذا حقق ما يلي:

- هو من الشكل النظامي الأول.
- كل الواصفات التي لا تشكل جزءاً من المفتاح الأساسي، تعتمد وظيفياً وبشكل كلي على المفتاح الأساسي.

التبعية الوظيفية (functional dependency):

تعتمد الواصفة B على الواصفة A وظيفياً إذا كانت كل قيمة لـ A تقابلها قيمة وحيدة لـ B، فمثلاً رقم الزبون يحدد اسمه وكل قيمة لـ c_id تتبع وظيفياً لـ c_name أن c_id لذلك يقال أن c_name لذلك يقال أن c_name).

B تعتمد وظيفياً بشكل كلي على مفتاح مركب (fully functional dependency)، إذا كان المفتاح المركب يحدد B، و B لا تعتمد وظيفياً على جزء منه.

تكتب علاقات التبعية بين الواصفات على الشكل التالي:

 $C_{id} \rightarrow c_{name}$

 $P_num \to p_address, owner_num, \ owner$

وضع الجداول في الشكل النظامي الثاني يبدأ بتحديد جميع علاقات التبعية بين الواصفات، ومن ثم تجزيء الجداول (decomposition) بشكل يضمن اعتماد جميع الواصفات التي لا تشكل جزءاً من المفتاح الأساسي كلياً وبشكل مباشر على المفتاح الأساسي.

يمكن الاستنتاج من التعريف السابق أن جميع الجداول ذات المفتاح البسيط (غير المركب) والتي تخضع لمعيار INF هي حتماً من الشكل النظامي الثاني.

بتطبيق ذلك على مثالنا نجد العلاقات التالية:

- 1. $C_{id} + p_{num} \rightarrow r_{start}$, r_{end} .
- **2.** C_id \rightarrow c_name.
- 3. P_num \rightarrow p_address, rent, owner_num, owner.
- **4.** Owner_num \rightarrow owner.
- 5. C_id + r_start \rightarrow p_num, p_address, r_end, r_end, rent, owner_num, owner.
- **6.** P_num + r_start \rightarrow c_id, c_name, r_end.

بمناقشة هذه العلاقات نجد ما يلي:

1. C_id + p_num \rightarrow r_start, r_end.

r_start, r_end تعتمد على (c_id, p_num) وبما أن (c_id, p_num) هو المفتاح المركب فذلك لا يناقض مع 2NF.

2. $C_{id} \rightarrow c_{name}$.

C_name تعتمد جزئياً على المفتاح الأساسي (partial functional dependency)، هذا يتناقض مع 2NF.

3. P_num \rightarrow p_address, rent, owner_num, owner.

هذه الحالة تشابه الحالة (2).

4. Owner_num \rightarrow owner.

owner_num تعتمد على owner_num و owner_num تعتمد على المفتاح الأساسي، أي أن owner تعتمد على المفتاح الأساسي التعدّي (transitive dependency)، وهذا لا يتناقض مع شرط 2NF.

5. C_id + r_start \rightarrow p_num, p_address, r_end, r_end, rent, owner_num, owner.

 c_{num} , p_address, r_end, r_start, rent, owner, owner_num تحدد c_{num} تحدد c_{num} ولكن c_{num} ولكن c_{num} ممكن أن يكون مفتاح أساسي، وبما أننا لم نقرر بعد (نحن في delid) هي مفتاح مرشح ممكن أن يكون مفتاح أساسي، وبما أننا لم نقرر بعد (نحن في delid) معندة البيانات) فإن هذه العلاقة لا تخرق شرط c_{num}

6. P_num + r_start \rightarrow c_id, c_name, r_end.

هذه الحالة تشابه الحالة (5).

يبدأ تطبيق معيار 2NF بتجزيء الجدول إنطلاقاً من علاقة التبعية الكاملة (1): C_id + p_num \rightarrow r_start, r_end

C_id	P_num	C_name	P_address	R_start	R_end	Rent	O_num	Owner
01	Pr3	Jane	А	1-1-96	12-1-98	785	Po23	Jones
01	Pr22	Jane	В	2-1-98	3-30-00	1200	Po44	Jan
02	Pr17	Fred	С	2-1-88	1-11-90	1000	Po32	Jill
03	Pr32	Ed	D	6-1-90	3-1-95	950	Po32	Jill
03	Pr22	Ed	В	4-1-00	null	1200	Po44	Jan

Rental

C_id	P_num	R_start	R_end
01	Pr3	1-1-96	12-1-98
01	Pr22	2-1-98	3-30-00
02	Pr17	2-1-88	1-11-90
03	Pr32	6-1-90	3-1-95
03	Pr22	4-1-00	Null

Customer

C_id	C_name
01	Jane
02	Fred
03	Ed

Property

P_num	P_address	rent	O_num	Owner
Pr3	A	785	Po23	Jones
Pr22	В	1200	Po44	Jan
Pr17	С	1000	Po32	Jill
Pr32	D	950	Po32	Jill

تحل جداول الشكل النظامي الثاني إشكاليات الحذف والإضافة ولا تحل إشكالية التعديل، فمثلاً إذا أردنا تعديل بيانات أحد المالكين فيجب تعديل مجموعة من التسجيلات مما يتفق مع إشكالية التعديل سابقة الذكر.

3NF) الشكل النظامي الثالث (3NF)

يكون الجدول من الشكل النظامي الثالث إذا حقق ما يلي:

- من الشكل النظامي الثاني.
- لا توجد فيه واصفات لا تشكل جزءاً من المفتاح الأساسي وتعتمد بالتعدّي على المفتاح الأساسي.

لا تحوي الجداول Rental و Customer و Customer واصفات تعتمد بالتعدي على المفتاح الأساسي، بينما تعتمد الواصفة owner في الجدول P_num في الجدول Ppoperty على owner_num الذي يعتمد بدوره على Ppoperty الشكل النظامي الثالث يجب تجزيئه إلى جدولين كما يلي:

Property

P_num	P_address	rent	O_num	Owner
Pr3	А	785	Po23	Jones
Pr22	В	1200	Po44	Jan
Pr17	С	1000	Po32	Jill
Pr32	D	950	Po32	Jill

Property

P_num	P_address	rent	O_num
Pr3	A	785	Po23
Pr22	В	1200	Po44
Pr17	С	1000	Po32
Pr32	D	950	Po32

Owner

O_num	Owner
Po23	Jones
Po44	Jan
Po32	Jill

الشكل النظامي الثالث يحل إشكاليات التعديل والحذف والإضافة، بالإضافة إلى حل مشكلة تكرار البيانات، وعندما يطلب تنظيم قاعدة بيانات فيقصد بذلك عادةً وضعها في الشكل النظامي الثالث.

4. تدریبات

- في الشكل النظامي الأول، تقاطع السطر مع العمود هو قيمة غير قابلة للتجزيء.
 - 1. صح
 - 2. خطأ

الإجابة: صح

- يمكن الإنتقال إلى الشكل النظامي الثالث من الشكل الأول دون المرور بالشكل النظامي الثاني.
 - 1. صح
 - 2. خطأ

الإجابة: خطأ

- في حال كان المفتاح الأساسي للجدول بسيط (غير مركب)، فلا فرق بين التبعية الوظيفية والتبعية الوظيفية الكلية.
 - 1. صح
 - 2. خطأ

الإجابة: صح

- أي من العبارات التالية تصف الشكل النظامي الثالث (اختر 2 من الإجابات)
 - 1. هو حتماً من الشكل النظامي الأول والثاني.
 - 2. المفتاح الأساسى للجدول هو حتماً مفتاح بسيط (غير مركب).
 - 3. لا توجد في الجدول علاقات تبعية بالتعدي.
- 4. الشكل النظامي الثالث يحل اشكالية الإدخال والحذف ولا يحل اشكالية التعديل.

الإجابة: (1-3)

- في حال الشكل النظامي الأول، كل خلية هي وحيدة القيمة (لا يمكن أن يكون محتوى الخلية هو مصفوفة من القيم).
 - 1. صح
 - 2. خطأ

الإجابة: صح

- في حال الشكل النظامي الأول، المدخلات في عمود هي من نفس النمط حتماً.
 - 1. صح
 - 2. خطأ

الإجابة: صح