مفاهيم قواعد البيانات

Database Concepts and Design

المستوى : الرابع

رمزالمقرر : ٢٢٣ حسب

المتطلبات السابقة : ١٢١ حسب

طبيعة المقرر : ساعتين نظري +ساعتين عملي

المرجع : أصول نظم قواعد البيانات –الجزء الأول

تأليف : أ.د رامز المصري / أ.د شامكانت نافاث

ترجمة د.م خالد ناصر السيد

استاذة المادة

م/ليندا البدري

الاعتمادات الوظيفية والتطبيع لقواعد البيانات العلائقية

(Functional dependencies and normalization for relational databases):

م/ ليندا البدري

المعلومات الزائدة في الصفوف العلاقية وشذوذ التحديث

- أحد أهداف تصميم المخطط هو إنقاص حيز التخزين الذي تشغلة العلاقات القاعدية في مخططات علاقات له تأثير هام على حيز التخزين .
- مثال : قارني بين الحيز المستخدم من قبل العلاقتين Department ، و Employee في الشكل رقم (١) مع الحيز المستخدم من قبل العلاقة DMP_DEPT في الشكل رقم (٢) وهو نتيجة تطبيق عملية (normal join) بين العلاقتين Department ، و Employee .

شكل (١)

EMPLOYEE

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

DEPARTMENT

Dname	Dnumber	Dmgr_ssn
Research	5	333445555
Administration	4	987654321
Headquarters	1	888665555

WORKS_ON

Ssn	Pnumber	Hours
123456789	1	32.5
123456789	2	7.5
666884444	3	40.0
453453453	1	20.0
453453453	2	20.0
333445555	2	10.0
333445555	3	10.0
333445555	10	10.0
333445555	20	10.0
999887777	30	30.0
999887777	10	10.0
987987987	10	35.0
987987987	30	5.0
987654321	30	20.0
987654321	20	15.0
[illegible]	20	Null

PROJECT

Pname	Pnumber	Plocation	Dnum
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newbenefits	30	Stafford	4

شكل (٢)

Redundancy

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	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

بالنظر الى الجداول السابقة نلاحظ الاتي :

- في العلاقة EMP_DEPT قيم الخواص (DMGRSSN وDNAME و DNUMBER) لقسم معين تتكرر مع كل موظف . في المقابل تظهر هذا المعلومات مرة واحدة فقط لكل قسم في العلاقة DEPARTMENT .
- في العلاقة EMPLOYEE يعاد رقم القسم لكل موظف يعمل لذلك القسم .
- نفس التعليقات يمكن تطبيقها على العلاقة EMP_PROJ في الشكل (٣) التي تضيف للعلاقة WORKS_ON خصائص من EMPLOYEEو PROJECT .

شكل (٣)

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
333445555	20	10.0	Wong, Franklin T.	Reorganization	Houston
999887777	30	30.0	Zelaya, Alicia J.	Newbenefits	Stafford
999887777	10	10.0	Zelaya, Alicia J.	Computerization	Stafford
987987987	10	35.0	Jabbar, Ahmad V.	Computerization	Stafford
987987987	30	5.0	Jabbar, Ahmad V.	Newbenefits	Stafford
987654321	30	20.0	Wallace, Jennifer S.	Newbenefits	Stafford
987654321	20	15.0	Wallace, Jennifer S.	Reorganization	Houston
888665555	20	Null	Borg, James E.	Reorganization	Houston

مشكلة شذوذ التحديث

- تصنف مشكلة شذوذ التحديث الى : (شذوذ إدخال ، شذوذ حذف ، شذوذ تحديث) .
- شذوذ الإدخال : يمكن تمييزه الى نوعين نوضحهما بالأمثلة التالية : ١- لادخال صف لموظف جديد في EMP_DEPT يجب أن يتضمن إما قيم الخصائص للقسم الذي يعمل به أو قيم خالية(اذا كان الموظف لا يعمل بقسم بعد).

٢- من الصعب أن ندخل قسماً جديداً لم يلحق به أي موظف بعد في العلاقة EMP_DEPT لانه في هذه الحالة نضطر الى وضع قيم خالية في خصائص الموظف (SSN هو المفتاح الاساسي)

- شذوذ الحذف : إذا حذفنا صف علائقي لموظف من EMP_DEPT وكان هذا الموظف آخر موظف يعمل في ذلك القسم ، فان المعلومات التي تخص ذلك القسم يتم فقدها من قاعدة البيانات .
- شذوذ التعديل : في العلاقة EMP_DEPT إذا غيرنا قيمة خاصية من خصائص قسم معين يجب تعديل الصفوف العلائقية لجميع الموظفين العاملين في ذلك القسم وإلا ستصبح قاعدة البيانات غير متجانسة .

لاحظي ان هذه المشاكل غير موجودة في الجداول في الشكل(١) .

الإعتمادات الوظيفية (الدالية)
Functional Dependencies

- الإعتماد الوظيفي (الدالي) : هو قيد بين مجموعتين من الخصائص من قاعدة البيانات .
- تعريف الإعتماد الوظيفي (الدالي) : ويرمز له بالرمز X→Y بين مجموعتين من الخصائص X , Y وهما مجموعتان جزئيتان من R ، يحدد قيداً على الصفوف العلائقية الممكنة التي يمكن أن تشكل حالة علاقة r من R .

وهذا يعني أن قيم العنصر Y في الصف العلائقي في r تعتمد على أو تحسب باستخدام قيم العنصر X . ونقول ايضاً أنه يوجد اعتماد وظيفي من X الى Y ، أو أن Y تعتمد وظيفياً على X (قيمة X تحدد أو تحسب قيمة Y) .

أمثلة للاعتمادات الوظيفية (FDs)

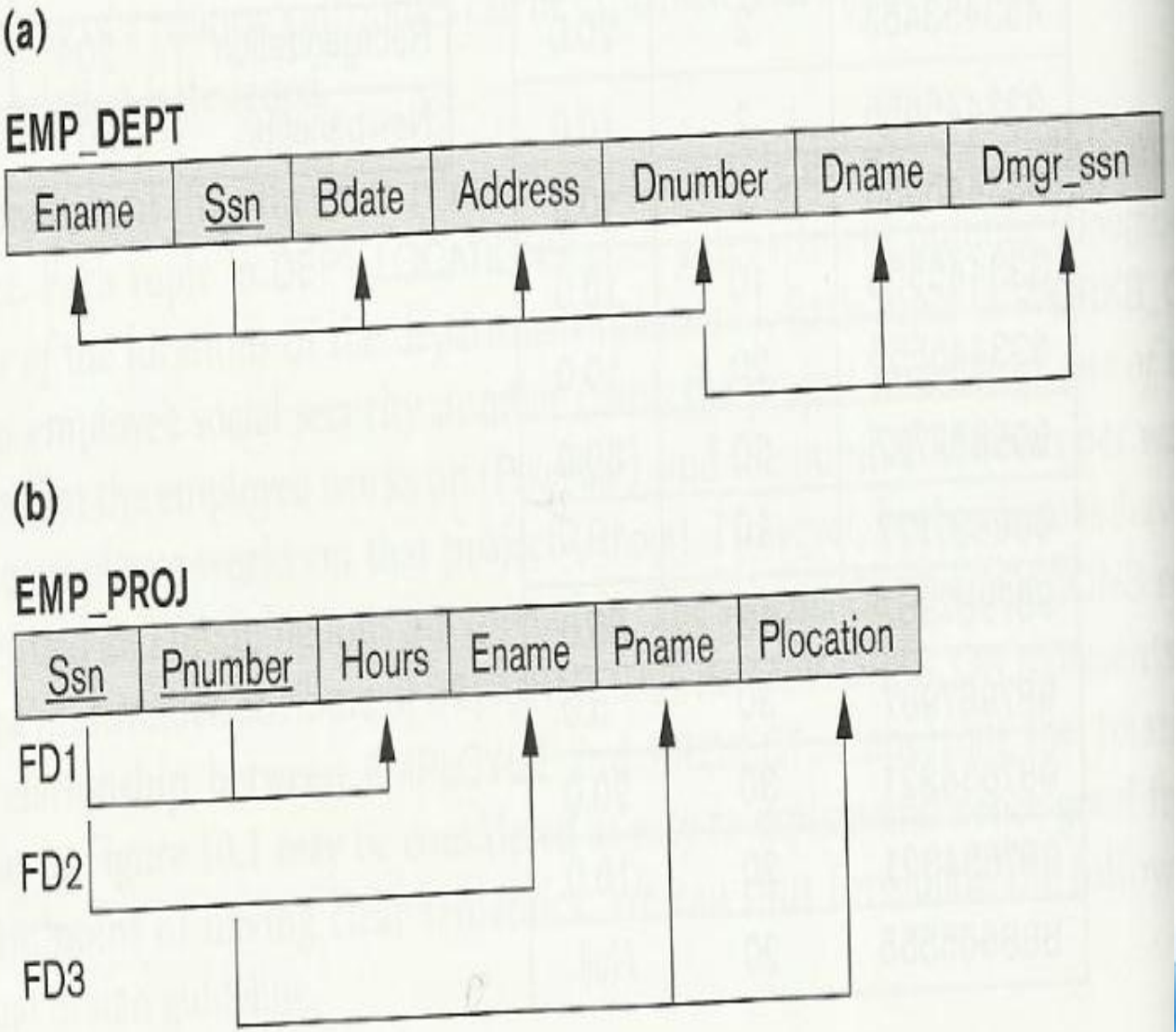
- بالنظر الى مخطط العلاقة EMP_PROJ في الشكل التالي (b) ومن الخصائص نستطيع أن نعرف الإعتمادات الوظيفية التالية :

A. Ssn→ Ename .

B. Pnumber → {Pname,Plocation} .

C. {SSN , Pnumber} → Hours .

Figure 10.3
Two relation schemas suffering from update anomalies.
(a) EMP_DEPT and
(b) EMP_PROJ.



- يعتبر الإعتماد الوظيفي صفة لمخطط العلاقة وليس لحالة معينة للعلاقة (FD لايمكن استخلاصه من حالة معينة).

مثال: الشكل التالي يعرض مخطط العلاقة teach .

- يمكن من النظرة الاولي أن نعتقد ان text → Course بالرغم من انه يمكن ان نكتشف فيما بعد قيم قد تؤدي الى بطلان الاعتماد الوظيفي .
- الاعتمادية الوظيفية Teacher → Course غير صحيحة (لماذا؟؟؟؟؟؟) .

TEACH

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

Figure 10.7
A relation state of TEACH with a *possible* functional dependency TEXT → COURSE. However, TEACHER → COURSE is ruled out.

تطبيع العلاقات (Normalization of relations)

- عملية التطبيع كما إقترحها كودد(١٩٧٢م) أولاً : تأخذ مخطط العلاقة خلال سلسة من الإختبارات لإثبات ما إذا كانت تتفق مع شكل طبيعي من الإشكال الطبيعية مع تفكيك العلاقات إذا لزم الأمر.
- في البداية إقترح كودد ثلاثة أشكال طبيعية هي : الشكل الطبيعي الأول والثاني و الثالث . ثم أُقْترح تعريف أقوي للشكل الطبيعي الثالث من قبل كودد وبويس .ثم تم أقتراح الشكل الطبيعي الرابع (4NF) ، و الشكل الطبيعي الخامس (5NF)

- يمكن النظر الى تطبيع البيانات كعملية تحليل مخططات العلاقات استنادا على إعتمادات البيانات بهم والمفاتيح الأساسية.
- اختبارات الشكل الطبيعي يتم تفكيكها الي مخططات علاقات أصغر تلبي الاختبارات ومن ثم تحتوي على الصفات المرغوبة.
- إجراء التطبيع يوفر لمصممي قاعدة البيانات مايلي :

- الإطار النظامي لتحليل مخططات العلاقات والذي يعتمد على مفاتيحهم الاساسية وعلى الإعتمادات الوظيفية بين خصائصهم .
- سلسلة من إختبارات الشكل الطبيعي يمكن تنفيذها على مخططات العلاقات وبذلك يمكن تطبيع قاعدة البيانات العلائقية الى أعلى درجة مطلوبة .

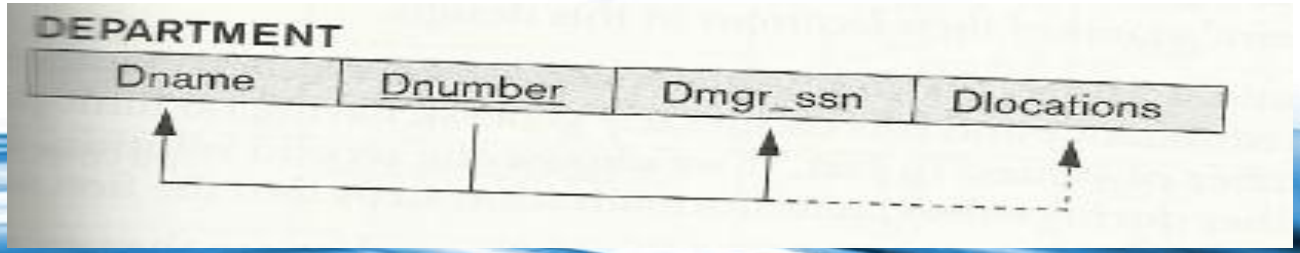
تعريف المفاتيح والخصائص المشاركة في المفاتيح
Definitions of keys and attributes participating in keys

- تعريف: المفتاح المميز لعلاقة (Superkey) $R=\{A_1, A_2, \ldots\ldots\ldots A_n\}$ هو مجموعة خصائص $S \subseteq R$ مع شرط أنه لايوجد صفيين علاقيين t_1 , t_2 في أي حالة علاقة قانونية r للعلاقة R سيكون لها $t_1 [S]= t_2 [S]$. المفتاح K هو المفتاح المميز Superkey فإذا حُذف أي خاصية من K سوف يسبب الا يكون مفتاحاً مميزاً بعد ذلك .
- في الشكل (١) (جدول Ssn (employee هو المفتاح المرشح الوحيد للعلاقة ، بينما {Ssn} ، {Ssn,Ename} ، {Ssn,Ename,Bdate} وغيرهم هي مفاتيح مميزة .

- واذا كان لمخطط علاقة اكثر من مفتاح ، أي منهم يسمى مفتاح مرشح ويكون الزامياً أن أحد المفاتيح المرشحة هو المفتاح الأساسي .
- في الشكل (١) (جدول Ssn (employee هو المفتاح المرشح الوحيد للعلاقة لذا يطلق عليه المفتاح الاساسي .
- تعريف : يطلق على خاصية لمخطط العلاقة R خاصية أساسية (Prime attribute) إذا كانت عضواً في مفتاح مرشح معين ل R.
- في الشكل (١) تكون كل من Ssn و Pnumber خصائص أساسية في Works_on بينما لا تعد الخصائص الأخرى أساسية .

First Normal Form الشكل الطبيعي الأول

- في هذا الشكل لايسمح بتعدد قيم الخصائص والخصائص المركبة واتحاداتهم . بكلمات أخرى أن قيم الخصائص المسموح بها في 1NF هي فقط قيم ذرية مفردة (غير قابلة للتجزئة).
- في مخطط العلاقة Department في الشكل (١) نجد أن المفتاح الأساسي هو Dnumber بفرض اننا توسعنا به ليشمل Dlocation كما هو مبين في الشكل التالي :



- هذه العلاقة ليست في 1NF لان Dlocation ليست خاصية ذرية كما هو مبين في الشكل التالي :

DEPARTMENT

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

- لانجاز 1NF لهذه العلاقة توجد ثلاث تقنيات :

١. نزيل الخاصية Dlocation التي تخالف 1NF ونصعها في علاقة منفصلة DEPT_LOCATIONS مع المفتاح الأساسي Dnumber .والمفتاح الأساسي لتلك العلاقة هو اتحاد {Dnumber , Dlocation} . وهذا يعني تفكيك العلاقة التي ليست في شكل 1NF الى علاقتين 1NF .

٢. نوسع المفتاح بحيث يوجد صف علائقي لكل موقع لقسم كما في الشكل التالي ، ويصبح المفتاح في هذه الحالة اتحاد {Dnumber , Dlocation} . وهذا الحل به عيب وهو ادخال التكرار في العلاقة .

(c)

DEPARTMENT

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

٣. إذا كان العدد الأقصى لقيم الخاصية معلوماً مثلاً إذا كان معلوماً انه يوجد ثلاث مواقع على الاكثر لكل قسم نستبدل خاصية بثلاث خصائص ذرية : Dlocation1,Dlocation2, Dlocation3 . وهذا الحل به عيب وهو أدخال فراغ إذا كان أغلب الأقسام لديها مواقع أقل من ثلاثة .

- الحل الأول يعد الحل الأفضل لانه لا يعاني من التكرار و هو عام تماماً ، وليس به حد أقصى لعدد القيم .
- لو أخترنا الحل الثاني سيتم تفكيكه فيما بعد أثناء خطوات تطبيع تالية الى الحل الأول .

- لا يسمح 1NF بالخصائص متعددة القيم التي هي نفسها مركبة (العلاقات المتداخلة كالعش).

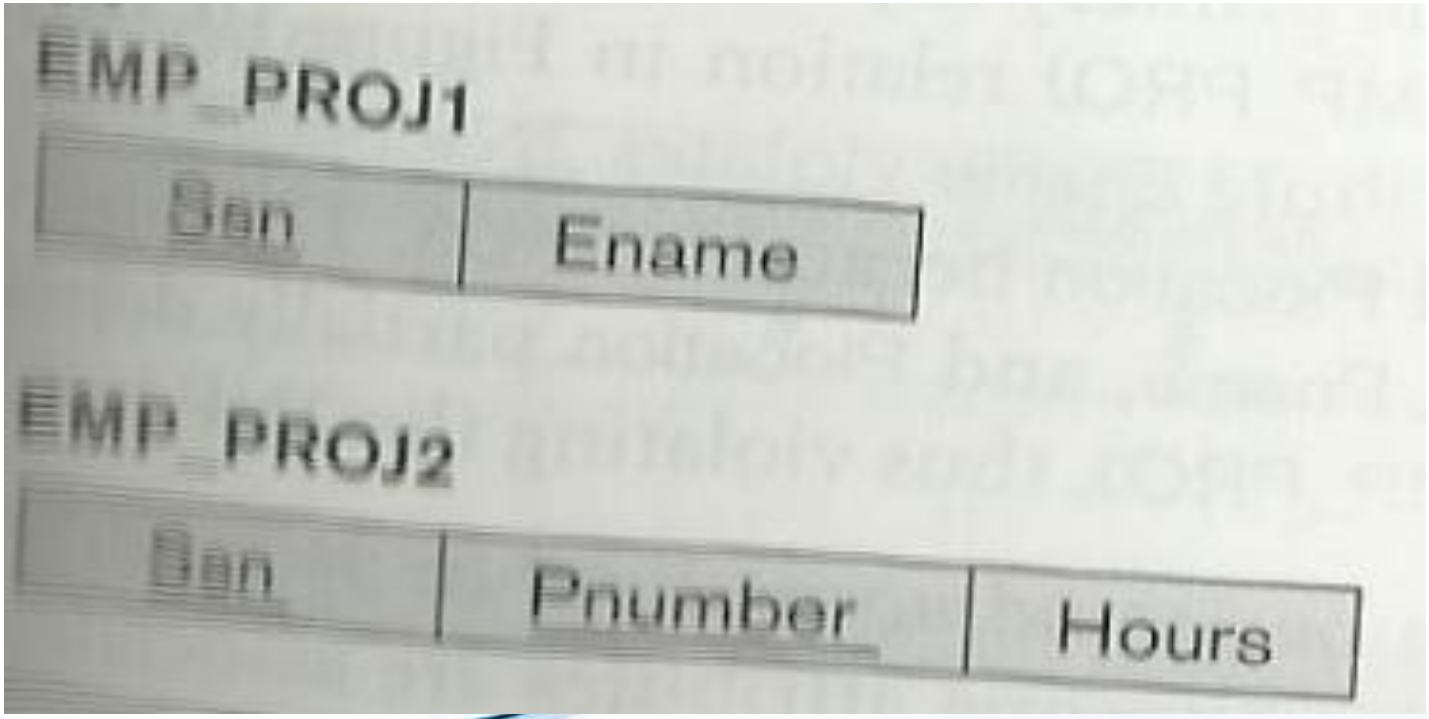
EMP_PROJ

Ssn	Ename	Projs	
		Pnumber	Hours

EMP_PROJ

Ssn	Ename	Pnumber	Hours
123456789	Smith, John B.	1	32.5
		2	7.5
666884444	Narayan, Ramesh K.	3	40.0
453453453	English, Joyce A.	1	20.0
		2	20.0
333445555	Wong, Franklin T.	2	10.0
		3	10.0
		10	10.0
		20	10.0
999887777	Zelaya, AliciaJ.	30	30.0
		10	10.0
987987987	Jabbar, Ahmad V.	10	35.0
		30	5.0
987654321	Wallace, Jennifer S.	30	20.0
		20	15.0
888665555	Borg, James E.	20	NULL

تفكيك العلاقة EMP_PROJ الى علاقات EMP_PROJ1, EMP_PROJ2 من درجة 1NF بإنتشار المفتاح الأساسي



- في حالة وجود أكثر من خاصية متعددة القيم في علاقة واحدة يجت التعامل معها بحرص وكمثال تأملي العلاقة التالية التي ليست في شكل 1NF :

Person (SS#,{CAR_LIC#},{PHONE#})

تمثل هذه العلاقة أن شخصاً يمتلك عدة سيارات ولديه عدة هواتف .

ملاحظة: الاقواس {} تعني تعدد القيم .

الطريقة السليمة للتعامل مع الخصائص متعددة القيم في Person هي تفكيكها الى علاقتين منفصلتين باستخدام الاستراتيجية التالية :

P1(SS#,CAR_LIC#),p2(SS#,PHONE#)

THE END