

Introduction

Levels at which we can discuss *goodness* of relation schemas
الخير من مخططات العلاقة

Logical (or conceptual) level (أو المفاهيمي)

Implementation (or physical storage) level

مستوى التنفيذ (أو التخزين المادي)

Approaches to database design:

مناهج تصميم قاعدة البيانات:

Bottom-up or top-down من أسفل إلى أعلى أو من أعلى إلى أسفل

Informal Design Guidelines for Relation Schemas

إرشادات التصميم غير الرسمية لمخططات العلاقة

Measures of quality مقاييس الجودة ■

- Making sure attribute semantics are clear

- التأكد من وضوح دلالات السمة

- Reducing redundant information in tuples

- تقليل المعلومات الزائدة في المجموعات

- Reducing NULL values in tuples

- تقليل القيم الفارغة في المجموعات

- Disallowing possibility of generating spurious

- tuples عدم السماح بإمكانية إنشاء مجموعات زائفة

Imparting Clear Semantics to Attributes in Relations

إضفاء دلالات واضحة على السمات في العلاقات

■ Semantics of a relation دلالات العلاقة

■ Meaning resulting from interpretation of attribute values in a tuple

■ المعنى الناتج عن تفسير قيم السمات في المجموعة

■ Easier to explain semantics of relation

■ أسهل في شرح دلالات العلاقة

■ Indicates better schema design

■ يشير إلى تصميم مخطط أفضل

Guideline 1

المبدأ التوجيهي ١

Design relation schema so that it is easy to explain its meaning

صمم مخطط العلاقة بحيث يسهل شرح معناها

Do not combine attributes from multiple entity types and relationship types into a single relation

لا تدمج السمات من أنواع كيانات وأنواع علاقات متعددة في علاقة واحدة

Example of violating Guideline 1: Figure 15.3

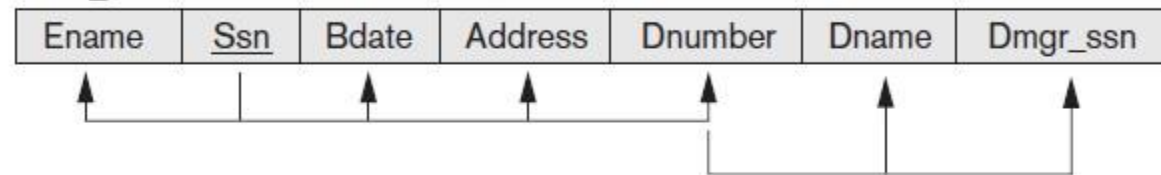
Guideline 1 (cont'd.)

Figure 15.3

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

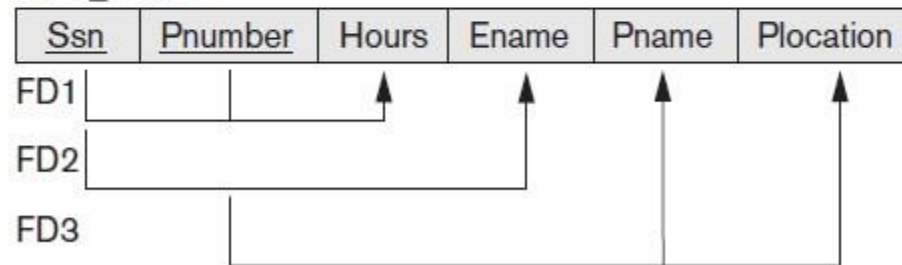
(a)

EMP_DEPT



(b)

EMP_PROJ



Redundant Information in Tuples and Update

معلومات زائدة في المجموعات وتحديث
الشذوذ

■ Grouping attributes into relation schemas

■ سمات التجميع في مخططات العلاقة

■ Significant effect on storage space

■ تأثير كبير على مساحة التخزين

■ Storing natural joins of base relations leads to **update anomalies**
الانحرافات يؤدي تخزين الصلات الطبيعية للعلاقات الأساسية إلى تحديث

■ Types of update anomalies: أنواع التحديثات الشاذة:

■ Insertion إدراج

■ Deletion حذف

■ Modification تعديل

Guideline 2

- Design base relation schemas so that no update anomalies are present in the relations
تصميم مخططات العلاقة الأساسية بحيث لا توجد حالات شاذة في التحديث في العلاقات
- If any anomalies are present: شذوذ:
في حالة وجود أي

- Note them clearly لاحظها بوضوح
- Make sure that the programs that update the database will operate correctly
- تأكد من أن البرامج التي تقوم بتحديث قاعدة البيانات ستعمل بشكل صحيح

NULL Values in Tuples

القيم الفارغة في المجموعات

May group many attributes together into a "fat" relation
"سمين" قد تجمع العديد من السمات معًا في علاقة

Can end up with many NULLs
من القيم الخالية يمكن أن ينتهي بالعديد

Problems with NULLs
مشاكل مع القيم الخالية

Wasted storage space
تخزين ضائعة مساحة

Problems understanding meaning

مشاكل فهم المعنى

Guideline 3

- Avoid placing attributes in a base relation whose values may frequently be NULL
- تجنب وضع السمات في علاقة أساسية قد تكون قيمها فارغة بشكل متكرر

- If NULLs are unavoidable:
- إذا كانت القيم الخالية لا مفر منها:

- Make sure that they apply in exceptional cases only, not to a majority of tuples
- تأكد من أنها تنطبق في حالات استثنائية فقط ، وليس على غالبية المجموعات

Generation of Spurious Tuples

Figure 15.5(a) ■

Relation schemas EMP_LOCS and ■
EMP_PROJ1

NATURAL JOIN انضمام طبيعي ■

Result produces many more tuples than the ■
original set of tuples in EMP_PROJ

ينتج عن النتيجة عدد من المجموعات أكثر من مجموعة ■
المجموعات الأصلية في EMP_PROJ

Called **spurious tuples** يسمى tuple الزائفة ■

Represent spurious information that is not valid ■

تمثل معلومات زائفة غير صحيحة ■

Guideline 4

Design relation schemas to be joined with equality conditions on attributes that are appropriately related ■

تصميم مخططات العلاقة ليتم ضمها مع شروط المساواة على السمات ذات الصلة بشكل مناسب ■

Guarantees that no spurious tuples are generated ■
يضمن عدم إنشاء مجموعات زائفة

Avoid relations that contain matching attributes that are not (foreign key, primary key) combinations ■

تجنب العلاقات التي تحتوي على سمات مطابقة ليست مجموعات (مفتاح خارجي ، مفتاح أساسي) ■

Summary and Discussion of Design Guidelines

ملخص ومناقشة إرشادات التصميم

Anomalies cause redundant work to be done ■
تؤدي الحالات الشاذة إلى عمل زائد عن الحاجة

Waste of storage space due to NULLs ■

هدر مساحة التخزين بسبب القيم الخالية ■

Difficulty of performing operations and joins ■

due to NULL values

بسبب قيم NULL
صعوبة تنفيذ العمليات والصلات

Generation of invalid and spurious data ■

during joins

الصلات

Functional Dependencies

Formal tool for analysis of relational schemas
أداة رسمية لتحليل المخططات العلائقية

Enables us to detect and describe some of the above-mentioned problems in precise terms
أعلاه بعبارات دقيقة
تمكننا من اكتشاف ووصف بعض المشاكل المذكورة

Theory of functional dependency

نظرية التبعية الوظيفية

Definition of Functional Dependency

Constraint between two sets of attributes ■
from the database

■ القيد بين مجموعتين من السمات من قاعدة البيانات

Definition. A functional dependency, denoted by $X \rightarrow Y$, between two sets of attributes X and Y that are subsets of R specifies a *constraint* on the possible tuples that can form a relation state r of R . The constraint is that, for any two tuples t_1 and t_2 in r that have $t_1[X] = t_2[X]$, they must also have $t_1[Y] = t_2[Y]$.

Property of semantics or meaning of the ■
attributes خاصية الدلالات أو معنى السمات

■ Legal relation states تتص العلاقة القانونية

■ Satisfy the functional dependency constraints

■ تلبية قيود التبعية الوظيفية

Definition of Functional Dependency (cont'd.)

- Given a populated relation نظرا لعلاقة مأهولة
- Cannot determine which FDs hold and which do not لا يمكن تحديد FDs التي تحمل وأيها لا
- Unless meaning of and relationships among attributes known ما لم يكن معنى العلاقات بين الصفات معروفة
- Can state that FD does not hold if there are tuples that show violation of such an FD
- يمكن أن يذكر أن FD لا يتم الاحتفاظ به في حالة وجود مجموعات تظهر انتهاكاً لـ FD

Normal Forms Based on Primary Keys

النماذج العادية على أساس المفاتيح الأساسية
Normalization process عملية التطبيع

Approaches for relational schema design

مناهج لتصميم المخطط العلائقي

- Perform a conceptual schema design using a conceptual model then map conceptual design into a set of relations

- قم بتنفيذ تصميم مخطط مفاهيمي باستخدام نموذج مفاهيمي ثم قم بتعيين التصميم المفاهيمي في مجموعة من العلاقات

- Design relations based on ex derived from existing implementation of files or forms or reports

- علاقات التصميم على أساس المعرفة الخارجية المستمدة من التنفيذ الحالي للملفات أو النماذج أو التقارير

- ternal knowledge المعرفة الأبدية

Normalization of Relations

تطبيع العلاقات

- Takes a relation schema through a series of tests يأخذ مخطط العلاقة من خلال سلسلة من الاختبارات
- Certify whether it satisfies a certain normal form
- تحقق مما إذا كانت تلبي شكلاً عادياً معيناً
- Proceeds in a top-down fashion العائدات بأسلوب تنازلي
- **Normal form tests** اختبارات الشكل العادي

Definition. The normal form of a relation refers to the highest normal form condition that it meets, and hence indicates the degree to which it has been normalized.

Normalization of Relations (cont'd.)

Properties that the relational schemas should have: ■
عليها المخططات العلائقية:

Nonadditive join property ■
خاصية الانضمام غير الإضافية

Extremely critical •
بالغة الأهمية

Dependency preservation property ■

خاصية الحفاظ على التبعية ■

Desirable but sometimes sacrificed for other factors •

مرغوب فيه ولكن في بعض الأحيان يتم التضحية به لعوامل أخرى •

Practical Use of Normal Forms

الاستخدام العملي للأشكال العادية

■ Normalization carried out in practice

■ نفذت التطبيع في الممارسة العملية

■ Resulting designs are of high quality and meet the desirable properties stated previously

■ التصميم الناتجة عالية الجودة وتفي بالخصائص المرغوبة المذكورة سابقاً

■ Pays particular attention to normalization only up to 3NF, BCNF, or at most 4NF

■ يولي اهتماماً خاصاً للتطبيع فقط حتى 3NF أو BCNF أو 4NF على الأكثر

■ Do not need to normalize to the highest possible normal form

■ لا تحتاج إلى التطبيع إلى أعلى شكل عادي ممكن

Definition. Denormalization is the process of storing the join of higher normal form relations as a base relation, which is in a lower normal form.

Definitions of Keys and Attributes Participating in Keys

تعريفات المفاتيح والسمات المشاركة في المفاتيح

Definition of **superkey** and **key** ■

Candidate key ■ مفتاح مرشح

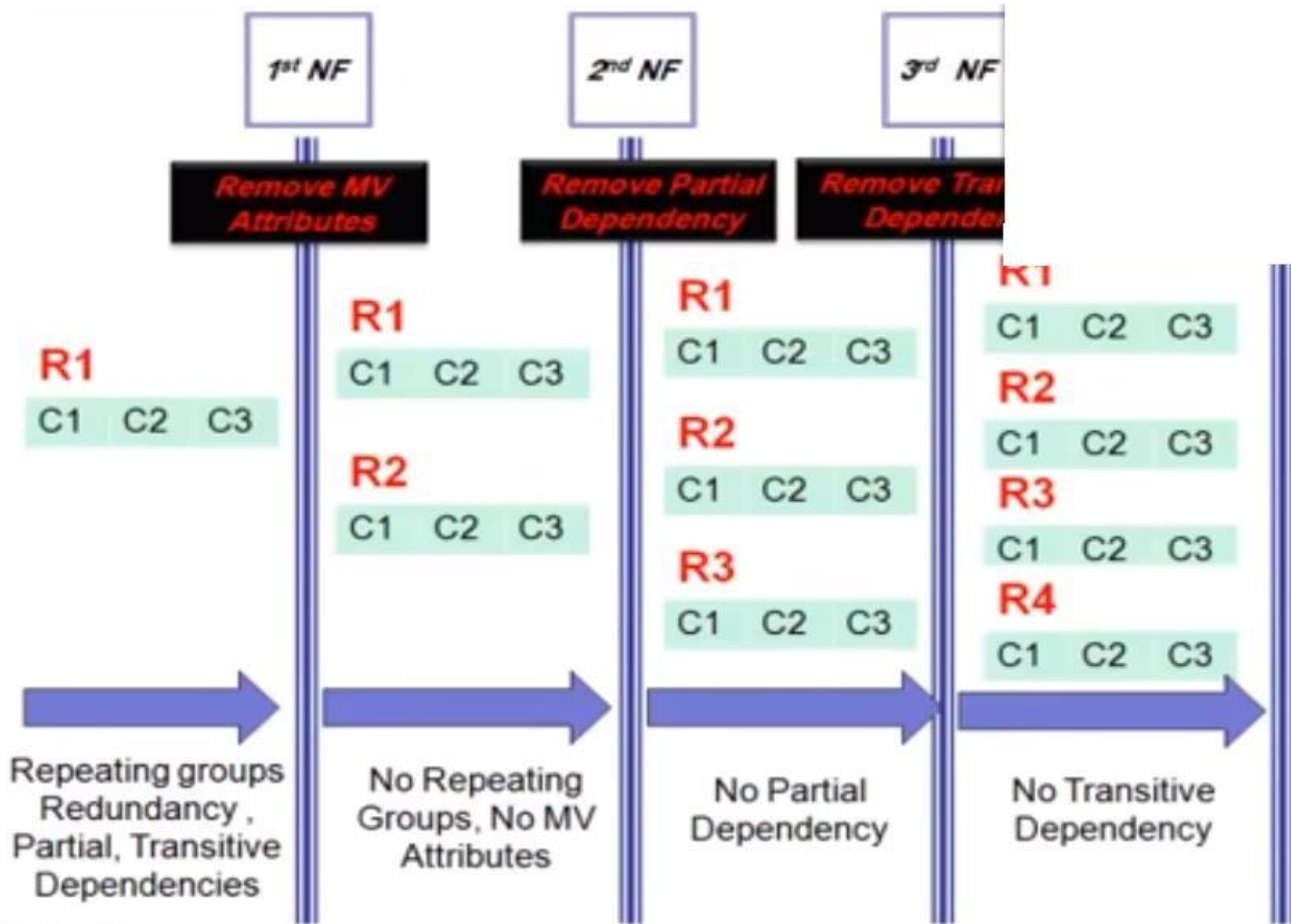
If more than one key in a relation schema ■

■ إذا كان هناك أكثر من مفتاح في مخطط العلاقة

One is **primary key** •

Others are **secondary keys** •

Definition. An attribute of relation schema R is called a **prime attribute** of R if it is a member of *some candidate key* of R . An attribute is called **nonprime** if it is not a prime attribute—that is, if it is not a member of any candidate key.



أول نموذج عادي First Normal Form

Part of the formal definition of a relation in the basic (flat) relational model ■

جزء من التعريف الرسمي للعلاقة في النموذج العلائقي الأساسي (المسطح) ■

Only attribute values permitted are single ■

فقط قيم السمات المسموح بها هي منفردة ■

atomic (or indivisible) values

Techniques to achieve first normal form عادي ■
تقنيات لتحقيق أول شكل

Remove attribute and place in separate relation ■

إزالة السمة والمكان في علاقة منفصلة ■

Expand the key قم بتوسيع المفتاح ■

Use several atomic attributes استخدم عدة سمات ذرية ■

First Normal Form (cont'd.)

■ Does not allow **nested relations** لا يسمح بالعلاقات المتداخلة

■ Each tuple can have a relation within it

■ يمكن أن يكون لكل مجموعة علاقة بداخلها

■ To change to 1NF: للتغيير إلى 1NF:

■ Remove nested relation attributes into a new

relation قم بإزالة سمات العلاقة المتداخلة في علاقة جديدة


■ Propagate the primary key into it انشر المفتاح الأساسي فيه

■ **Unnest** relation into a set of 1NF relations

■ علاقة غير مترابطة في مجموعة من العلاقات 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

Figure 15.9

Normalization into 1NF. (a) A relation schema that is not in 1NF. (b) Sample state of relation DEPARTMENT. (c) 1NF version of the same relation with redundancy.

Second Normal Form

Based on concept of **full functional dependency** ■

على أساس مفهوم التبعية الوظيفية الكاملة

Versus **partial dependency** ■ مقابل التبعية الجزئية

Definition. A relation schema R is in 2NF if every nonprime attribute A in R is *fully functionally dependent* on the primary key of R .

Second normalize into a number of 2NF ■

relations التطبيع الثاني في عدد من العلاقات NF₂

Nonprime attributes are associated only with ■
part of primary key on which they are fully
functionally dependent

■ السمات غير الممتازة مرتبطة فقط بجزء من المفتاح الأساسي
الذي تعتمد عليه وظيفيًا بشكل كامل

Third Normal Form

- Based on concept of transitive dependency ■
- على أساس مفهوم التبعية متعدية ■

Definition. According to Codd's original definition, a relation schema R is in 3NF if it satisfies 2NF *and* no nonprime attribute of R is transitively dependent on the primary key.

FD إشكالية Problematic FD ■

- Left-hand side is part of primary key ■
- الجانب الأيسر جزء من المفتاح الأساسي ■
- Left-hand side is a nonkey attribute ■
- الجانب الأيسر سمة nonkey ■

General Definitions of Second and Third Normal Forms

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).

General Definitions of Second and Third Normal Forms (cont'd.)

■ **Prime attribute السمة الأولية**

Part of any candidate key will be considered as prime ■
prime سيتم اعتبار جزء من أي مفتاح مرشح على أنه أولي

■ Consider partial, full functional, and transitive dependencies with respect to all candidate keys of a relation

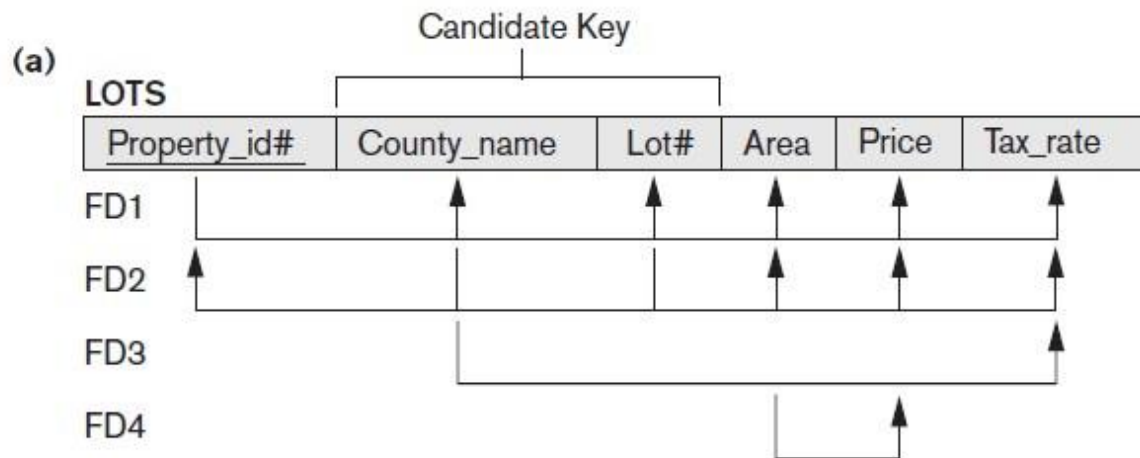
■ ضع في اعتبارك التبعية الجزئية والكاملة والوظيفية والمتعددة فيما يتعلق بجميع المفاتيح المرشحة لعلاقة ما

General Definition of Second Normal Form

Definition. A relation schema R is in **second normal form (2NF)** if every non-prime attribute A in R is not partially dependent on *any* key of R .¹¹

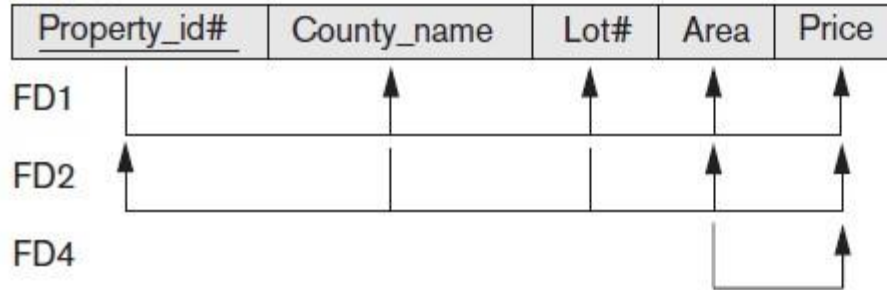
Figure 15.12

Normalization into 2NF and 3NF. (a) The LOTS relation with its functional dependencies FD1 through FD4. (b) Decomposing into the 2NF relations LOTS1 and LOTS2. (c) Decomposing LOTS1 into the 3NF relations LOTS1A and LOTS1B. (d) Summary of the progressive normalization of LOTS.

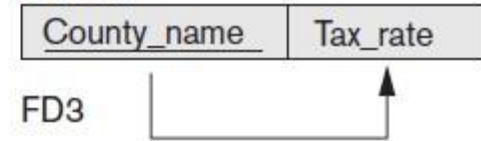


(b)

LOTS1

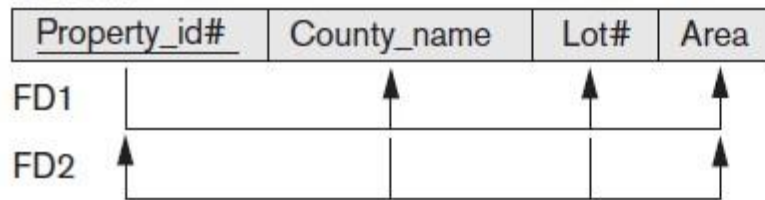


LOTS2

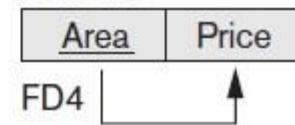


(c)

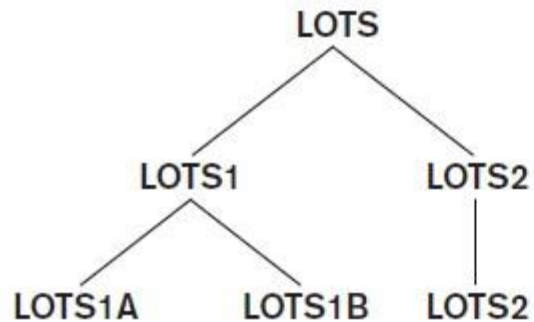
LOTS1A



LOTS1B



(d)



1NF

2NF

3NF

General Definition of Third Normal Form

Definition. A relation schema R is in **third normal form (3NF)** if, whenever a *nontrivial* functional dependency $X \rightarrow A$ holds in R , either (a) X is a superkey of R , or (b) A is a prime attribute of R .

Alternative Definition. A relation schema R is in 3NF if every nonprime attribute of R meets both of the following conditions:

- It is fully functionally dependent on every key of R .
- It is nontransitively dependent on every key of R .

Boyce-Codd Normal Form نموذج Boyce-Codd العادي

Every relation in BCNF is also in 3NF ■

كل علاقة في BCNF هي أيضًا في 3NF ■

Relation in 3NF is not necessarily in BCNF ■

العلاقة في 3NF ليست بالضرورة في BCNF ■

Definition. A relation schema R is in BCNF if whenever a *nontrivial* functional dependency $X \rightarrow A$ holds in R , then X is a superkey of R .

Difference: ■

Condition which allows A to be prime is absent from ■

BCNF الشرط الذي يسمح لـ A أن يكون أوليًا غائب من BCNF

Most relation schemas that are in 3NF are also in BCNF ■

معظم مخططات العلاقات الموجودة في 3NF موجودة أيضًا في BCNF ■

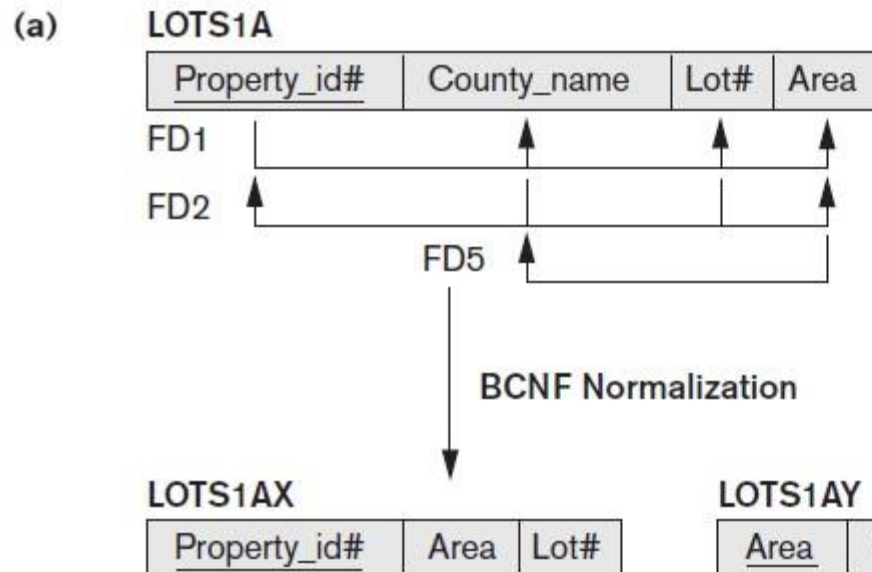
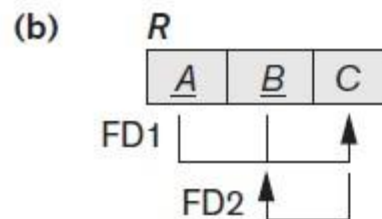


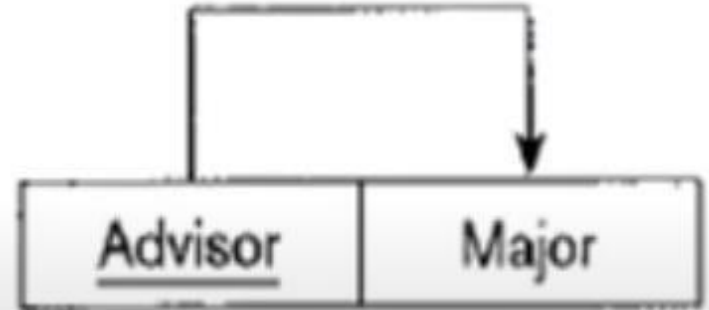
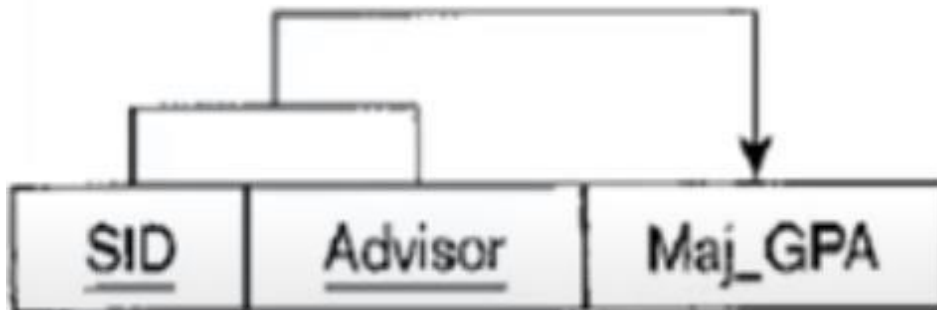
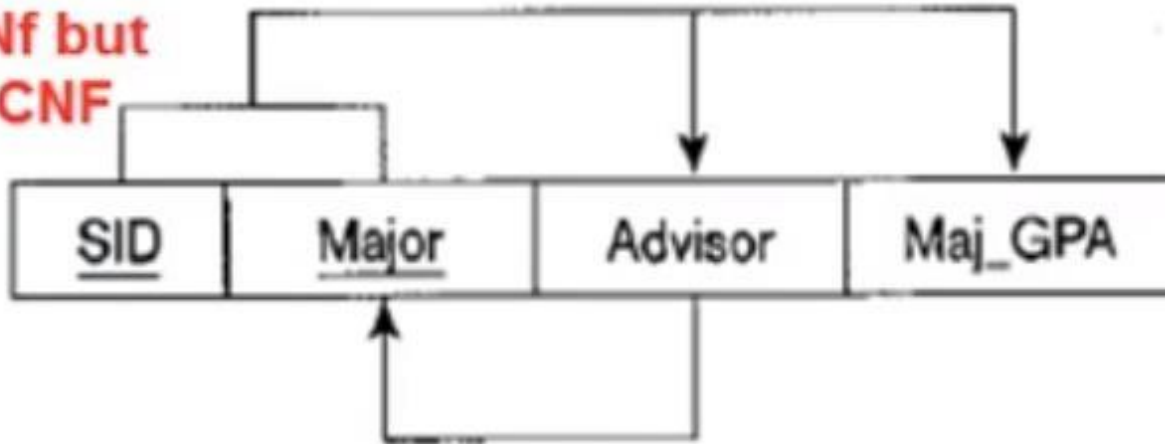
Figure 15.13

Boyce-Codd normal form. (a) BCNF normalization of LOTS1A with the functional dependency FD2 being lost in the decomposition. (b) A schematic relation with FDs; it is in 3NF, but not in BCNF.



**R in 3rd Nf but
not in BCNF**

R



Multivalued Dependency and Fourth Normal Form

التبعية متعددة القيم والنموذج الرابع

العادي الرابع

Multivalued dependency (MVD) ■

التبعية متعددة القيم (MVD) ■

Consequence of first normal form (1NF) ■

نتيجة النموذج العادي الأول (1NF) ■

Definition. A multivalued dependency $X \twoheadrightarrow Y$ specified on relation schema R , where X and Y are both subsets of R , specifies the following constraint on any relation state r of R : If two tuples t_1 and t_2 exist in r such that $t_1[X] = t_2[X]$, then two tuples t_3 and t_4 should also exist in r with the following properties,¹⁵ where we use Z to denote $(R - (X \cup Y))$:¹⁶

- $t_3[X] = t_4[X] = t_1[X] = t_2[X]$.
- $t_3[Y] = t_1[Y]$ and $t_4[Y] = t_2[Y]$.
- $t_3[Z] = t_2[Z]$ and $t_4[Z] = t_1[Z]$.

Multivalued Dependency and Fourth Normal Form (cont'd.)

- Relations containing nontrivial MVDs ■

- العلاقات التي تحتوي على MVDs غير بديهية ■

- All-key relations العلاقات الشاملة ■

- Fourth normal form (4NF) النموذج العادي الرابع (4NF) ■

- Violated when a relation has undesirable multivalued dependencies ■

- تنتهك عندما يكون للعلاقة تبعيات متعددة القيم غير مرغوب فيها ■

Definition. A relation schema R is in 4NF with respect to a set of dependencies F (that includes functional dependencies and multivalued dependencies) if, for every *nontrivial* multivalued dependency $X \twoheadrightarrow Y$ in F^{+17} X is a superkey for R .

Join Dependencies and Fifth Normal Form

انضم إلى التبعيات والنموذج العادي الخامس Form

- **Join dependency**
 - Multiway decomposition into fifth normal form (5NF)
 - التحلل متعدد الطرق إلى الشكل العادي الخامس (5NF)
 - Very peculiar semantic constraint
 - القيد الدلالي غريب جدا
 - Normalization into 5NF is very rarely done in practice
- ادرًا ما يتم إجراء التطبيع إلى 5NF
- في الممارسة العملية

Join Dependencies and Fifth Normal Form (cont'd.)

Definition. A join dependency (JD), denoted by $JD(R_1, R_2, \dots, R_n)$, specified on relation schema R , specifies a constraint on the states r of R . The constraint states that every legal state r of R should have a nonadditive join decomposition into R_1, R_2, \dots, R_n . Hence, for every such r we have

$$* (\pi_{R_1}(r), \pi_{R_2}(r), \dots, \pi_{R_n}(r)) = r$$

Definition. A relation schema R is in **fifth normal form (5NF)** (or **project-join normal form (PJNF)**) with respect to a set F of functional, multivalued, and join dependencies if, for every nontrivial join dependency $JD(R_1, R_2, \dots, R_n)$ in F^+ (that is, implied by F),¹⁸ every R_i is a superkey of R .