

Normal Forms for Relational Databases

关系数据库的范式

Normal Forms for Relational Databases

- criteria for a good database design (i.e., to resolve update anomalies)
- formalized by functional (or other) dependencies

Normal Forms for Relational Databases_(cont)

Normal Forms:

- 1NF, 2NF, 3NF (Codd 1972)
- Boyce-Codd NF (1974)
- Multivalued dependencies and 4NF (Zaniolo 1976 and Fagin 1977)
- Join dependencies (Rissanen 1977) and 5NF (Fagin 1979)

First Normal Form (1NF)

This simply means that attribute values are *atomic*, and is part of the definition of the relational model.

Atomic: multivalued attributes, 复合的 composite attributes, and their combinations are disallowed.

There is currently a lot of interests in non-first normal form databases, particularly those where an attribute value can be a table (nested relations).

Consider the table below, adapted from Desai.

First Normal Form (1NF) (cont)

Fac_Dept	Prof	Course Preferences	
		Course	Course_Dept
Comp Sci	Smith	353	Comp Sci
		379	Comp Sci
		221	Decision Sci
	Clark	353	Comp Sci
		351	Comp Sci
		379	Comp Sci
		456	Mathematics
Chemistry	Turner	353	Comp Sci
		456	Mathematics
		272	Chemsitry
Mathematics	Jameison	353	Comp Sci
		379	Comp Sci
		221	Decision Sci
		456	Mathematics
		469	Mathematics

This can be transformed into:

First Normal Form (1NF) (cont)

CRS_PREF			
Prof	Course	Fac_Dept	Crs_Dept
Smith	353	Comp Sci	Comp Sci
Smith	379	Comp Sci	Comp Sci
Smith	221	Comp Sci	Decision Sci
Clark	353	Comp Sci	Comp Sci
Clark	351	Comp Sci	Comp Sci
Clark	379	Comp Sci	Comp Sci
Clark	456	Comp Sci	Mathematics
Turner	353	Chemistry	Comp Sci
Turner	456	Chemistry	Mathematics
Turner	272	Chemistry	Chemistry
Jamieson	353	Mathematics	Comp Sci
Jamieson	379	Mathematics	Comp Sci
Jamieson	221	Mathematics	Decision Sci
Jamieson	456	Mathematics	Mathematics
Jamieson	469	Mathematics	Mathematics

First Normal Form (1NF) (cont)

The representation in the figure above has the following drawbacks:

- the fact that a given professor is in a given department may be repeated,
- the association between professor and department will not be recorded unless the professor has some course references,
- the fact that a given course is offered by a given department may be repeated,
- again, this is not recorded unless someone has a preference for the course.

所谓第一范式 (1NF) 是指数据库表的每一列都是不可分割的基本数据项, 同一列中不能有多个值, 即实体中的某个属性不能有多个值或者不能有重复的属性

First Normal Form (1NF) (cont) 无重复的列

Suppose the FD's for these attributes are

$$F = \{Prof \rightarrow Fac_Dept, Course \rightarrow Crs_Dept\}.$$

Notice that a superkey is just a set of attributes S such that

$$S \rightarrow \{Prof, Course, Fac_Dept, Crs_Dept\} \in F^+$$

Thus the only candidate key here is $\{Prof, Course\}$.

First Normal Form (1NF) (cont)

These problems arise because *Fac_Dept* depends only on *Prof* and not on *Course*, and similarly *Crs_Dept* depends only on *Course* and not on *Prof*.

We can recognize and avoid these problems using functional dependencies.

Second Normal Form (2NF) 属性完全依赖于主键

A *prime* attribute is one that is part of a candidate key. Other attributes are *non-prime*.

Definition: In an FD $X \rightarrow Y$, Y is *fully functionally dependent* on X if there is no $Z \subset X$ such that $Z \rightarrow Y$. Otherwise Y is *partially dependent* on X .



Proper Subset

Definition (*Second Normal Form*): A relation scheme is in second normal form (2NF) if all non-prime attributes are fully functionally dependent on the candidate keys.

A database scheme is in 2NF if all its relations are in 2NF.

第二范式 (2NF) 要求数据库表中的每个实例或行必须可以被唯一地区分。为实现区分通常需要为表加上一个列，以存储各个实例的唯一标识。这个唯一属性列被称为主关键字或主键、主码。

Second Normal Form (2NF) (cont)

Possible 2NF decomposition of the relation above is:

COURSE_PREF	
Prof	Course
Smith	353
Smith	379
Smith	221
Clark	353
Clark	351
Clark	379
Clark	456
Turner	353
Turner	456
Turner	272
Jamieson	353
Jamieson	379
Jamieson	221
Jamieson	456
Jamieson	469

COURSE	
Course	Dept
353	Comp Sci
379	Comp Sci
221	Decision Sci
351	Comp Sci
456	Mathematics
272	Chemistry
469	Mathematics

FACULTY	
Prof	Dept
Smith	Comp Sci
Clark	Comp Sci
Turner	Chemistry
Jamieson	Mathematics

Second Normal Form (2NF) (cont)

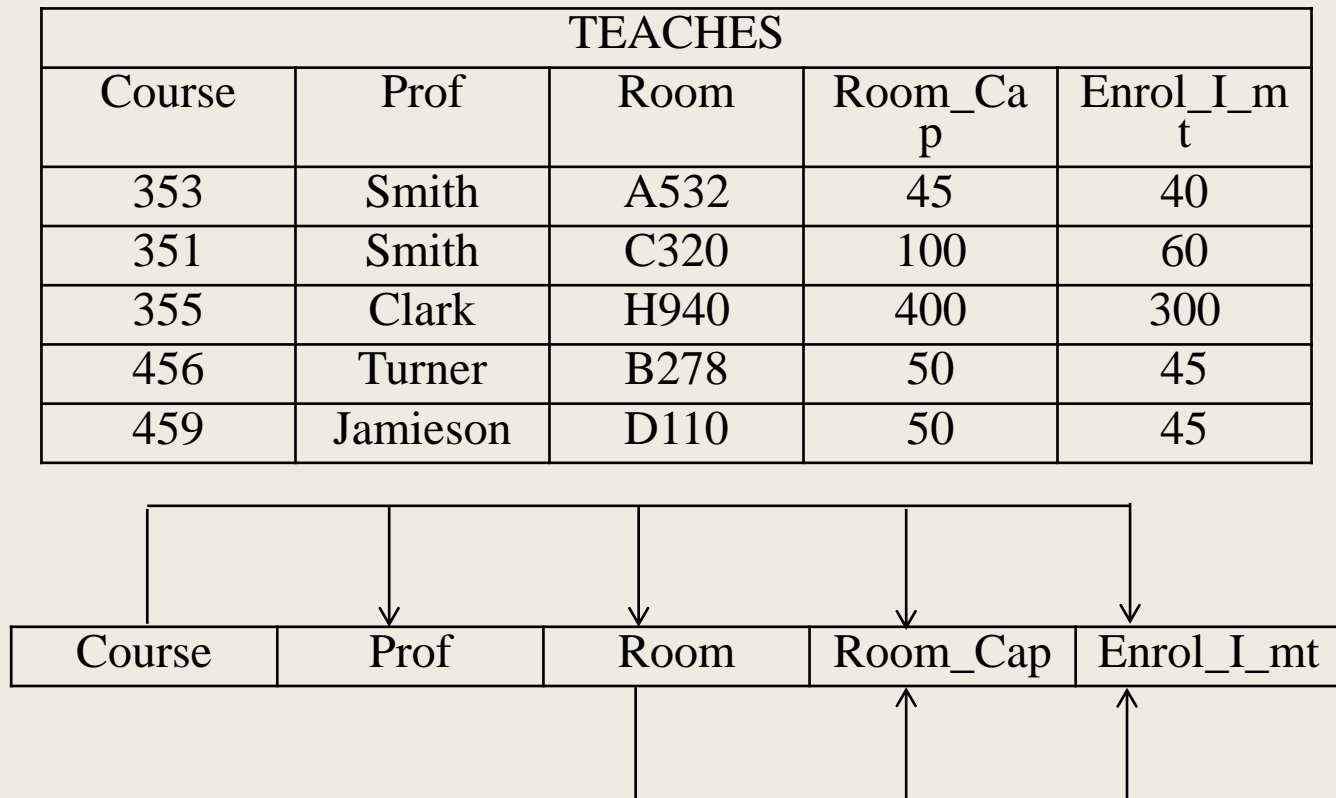
Question: What relational algebra expression recovers *CRS_PREF* from these relations?

Answer: Join

第二范式 (2NF) 要求实体的属性完全依赖于主关键字。所谓完全依赖是指不能存在仅依赖主关键字一部分的属性，如果存在，那么这个属性和主关键字的这一部分应该分离出来形成一个新的实体，新实体与原实体之间是一对多的关系。为实现区分通常需要为表加上一个列，以存储各个实例的唯一标识。

Second Normal Form (2NF) (cont)

2NF does not completely eliminate the kind of anomaly we saw before:



Second Normal Form (2NF) (cont)

This is in 2NF but:

If another course uses say Room A532, then the fact that A532 has *Room_Cap* of 45 and *Enrol_Lmt* of 40 will be stored twice.

If course 355 is deleted, then the fact that H940 has *Room_Cap* of 400 and *Enrol_Lmt* of 300 will be lost.

This we can also fix by adding further restrictions on functional dependencies.

Third Normal Form (3NF) 不依赖于其它非主属性;消除传递依赖

Definition: An FD $X \rightarrow Y$ is a transitive dependency if there is a Z that is not a subset of any key, such that $X \rightarrow Z$ and $Z \rightarrow Y$ and $Z \nrightarrow X$ hold.

The attributes of Y are transitively dependent on X .

第三范式 (3NF) 要求一个数据库表中不包含已在其它表中已包含的非主关键字信息。例如, 存在一个部门信息表, 其中每个部门有部门编号 (dept_id)、部门名称、部门简介等信息。那么在员工信息表中列出部门编号后就不能再将部门名称、部门简介等与部门有关的信息再加入员工信息表中。如果不存在部门信息表, 则根据第三范式 (3NF) 也应该构建它, 否则就会有大量的数据冗余。简而言之, 第三范式就是属性不依赖于其它非主属性。

e.g. $Room_Cap$ is transitively dependent on $\{Course\}$, since $\{Course\} \rightarrow \{Room\}$ and $\{Room\} \rightarrow \{Room_Cap\}$ hold, and $\{Room\}$ is not a subset of any key.

Third Normal Form (3NF) (cont)

Definition (Third Normal Form): A relation scheme is in *third normal form (3NF)* if for all non-trivial FD's of the form $X \rightarrow A$ that hold, either X is a superkey or A is a prime attribute.

Note: a FD $X \rightarrow Y$ is trivial iff Y is a subset of X .

Alternative definition: A relation scheme is in third normal form if every non-prime attribute is fully functionally dependent on the keys and not transitively dependent on any key.

A database scheme is in 3NF if all its relations are in 3NF.

Third Normal Form (3NF) (cont)

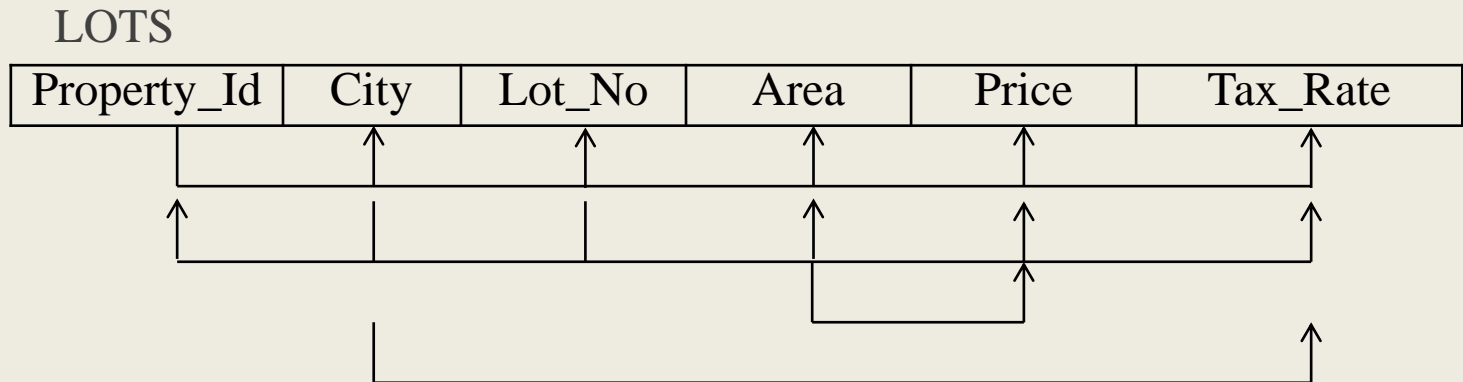
TEACHES can be decomposed into 3NF:

ROOM_DETAILS		
Room	Room_Cap	Enrol_I_mt
A532	45	40
C320	100	60
B278	50	45
D110	50	45
H940	400	300

COURSE_DETAILS		
Course	Prof	Room
353	Smith	A532
351	Smith	C320
456	Turner	B278
459	Jamieson	D110
355	Clark	H940

Third Normal Form (3NF) (cont)

Another example:

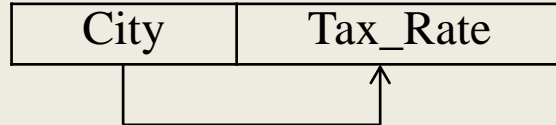


This is not in 2NF since $City \rightarrow Tax_Rate$, Tax_Rate is not prime, and $\{City, Lot_No\}$ is a key, making Tax_Rate partially dependent on a key.

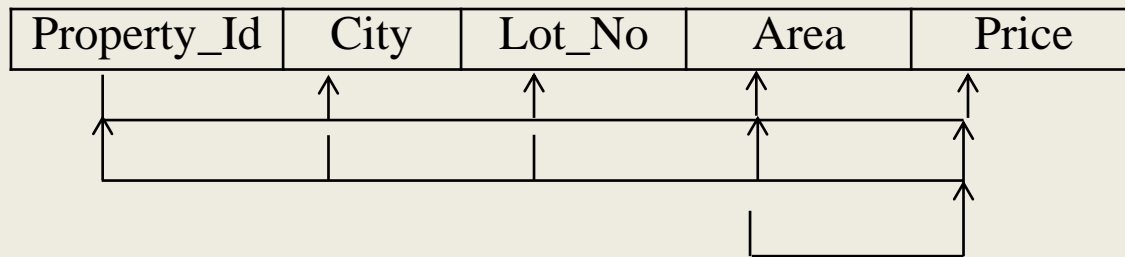
We could fix this:

Third Normal Form (3NF) (cont)

LOTS1



LOTS2



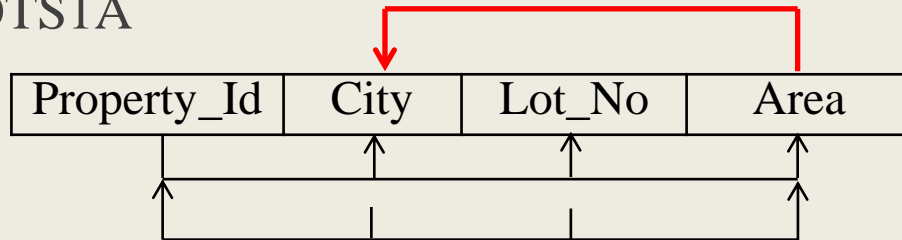
Now we have 2NF but not 3NF, since $Area \rightarrow Price$, $\{Area\}$ is not a superkey and Price is not prime.

Note: the transitive dependency : $Property_Id \rightarrow Area \rightarrow Price$.

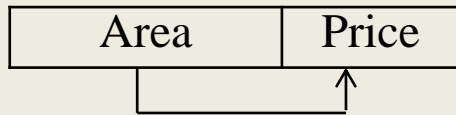
We could fix this too:

Third Normal Form (3NF) (cont)

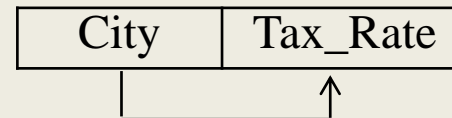
LOTS1A



LOTS1B



LOTS2



Suppose also that $Area \rightarrow City$. The relations schemes are still in 3NF since $City$ is a prime attribute. However, there can be anomalies, just as before. We need more restrictions still to fix these.

Boyce-Codd Normal Form (BCNF)

Definition (Boyce-Codd Normal Form):

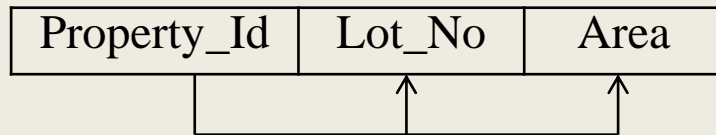
A relation scheme is in *Boyce-Codd Normal Form* (BCNF) if whenever $X \rightarrow A$ holds and $X \rightarrow A$ is non-trivial, X is a superkey.

A database scheme is in BCNF if all its relations are in BCNF.

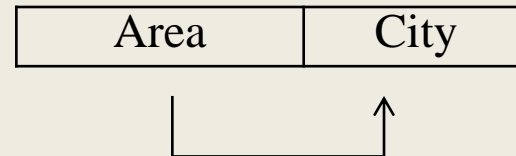
We can make our example into BCNF:

Boyce-Codd Normal Form (BCNF)_(cont)

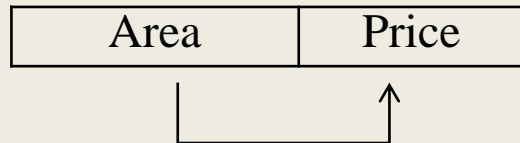
LOTS1AA



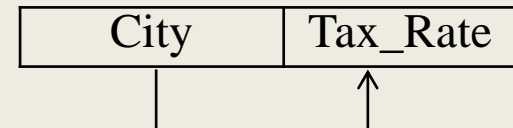
LOTS1AB



LOTS1B

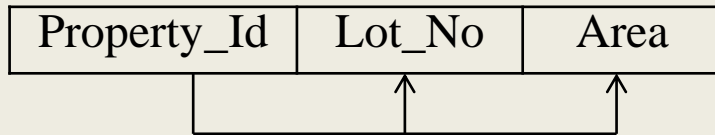


LOTS2

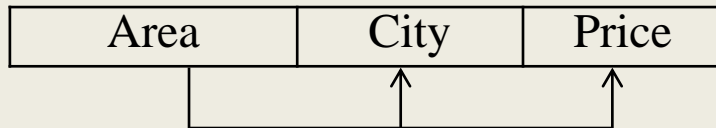


Boyce-Codd Normal Form (BCNF)_(cont)

LOTS1AA



LOTS1AB



LOTS2

