Week11_Handout

Database System
Schema+Design part6-7

‹#›

Week11_Handout

Database System-Schema+Design part6

‹#›

Schema Design and Refinement

- How do we obtain a good design?
- Functional dependencies & keys
- Desirable properties of schema refinement
- Boyce Codd Normal Form (BCNF)
- Third Normal Form (3NF) and 3NF Decomposition
- Fourth Normal Form (4NF)

Third Normal Form - Motivation

- There is one structure of FD's that causes trouble when we decompose.

  $AB \rightarrow C$ and $C \rightarrow B$ (Example: A = street address, B = city, C = zip.)

- There are two keys, $\{A,B\}$ and $\{A,C\}$.

  $C \rightarrow B$ is a BCNF violation, so we must decompose into AC, BC.

- The problem is that if we use AC and BC as our database schema, we cannot enforce the FD $AB \rightarrow C$ by checking FD's in these decomposed relations.

An Unenforceable FD

street	zip
545 Tech Sq.	02138
545 Tech Sq.	02139

city	zip
Cambridge	02138
Cambridge	02139

Join tuples with equal zip codes.

street	city	zip
545 Tech Sq.	Cambridge	02138
545 Tech Sq.	Cambridge	02139

Although no FD' s were violated in the decomposed relations,

FD street city → zip is violated by the database as a whole.

3NF Let' s Us Avoid This Problem

- 3rd Normal Form (3NF) modifies the BCNF condition so we do not have to decompose in this problem situation.
- An attribute is *prime* if it is a member of any key.
- $X \rightarrow A$ violates 3NF if and only if X is not a superkey, and also A is not prime.

Example

- In our problem situation with FD's

  $AB \rightarrow C$ and $C \rightarrow B$, we have keys AB and AC.
- Thus A, B, and C are each prime.
- Although $C \rightarrow B$ violates BCNF, it does not violate 3NF.

What 3NF and BCNF Give You

- There are two important properties of a decomposition:
  1. *Recovery* : it should be possible to project the original relations onto the decomposed schema, and then reconstruct the original.
  2. *Dependency Preservation* : it should be possible to check in the projected relations whether all the given FD's are satisfied.

3NF and BCNF, Continued

- We can get (1) with a BCNF decomposition.
  - Explanation needs to wait for relational algebra.
- We can get both (1) and (2) with a 3NF decomposition.
- But we can' t always get (1) and (2) with a BCNF decomposition.
  - street-city-zip is an example.

‹#›

保持函数依赖和无损连接的3NF模式分解算法

已知关系模式和函数依赖集合，保持FD和无损连接3NF模式分解的算法如下：

① 求极小函数依赖集S_{min}和关系模式的所有键（keys）；

② 在S_{min}中按函数依赖左部相同原则进行分组，每个组中的所有属性形成分解后的子关系模式，$R_1, R_2, ..., R_m$

③ 如果某个关系模式R_i的所有属性被另一个关系模式R_j所包含，删除关系模式R_i；

④ 判断是否有某个key出现在其中的一个关系模式中，如果出现则结束；如果没有出现，将任意一个key也作为子关系模式R_k加入到分解后的模式中。

‹#›

最小函数依赖集

如果函数依赖集合F满足如下条件，则称F为一个极小函数依赖集，也称为最小依赖及或最小覆盖。

（1）F中的任意函数依赖的右部仅含有一个属性；

（2）F中不存在这样的函数依赖$X \rightarrow A$，使得F与$F-\{X \rightarrow A\}$等价；

（3）F中不存在这样的函数依赖$X \rightarrow A$，X有真子集Z，

使得$F-\{X \rightarrow A\} \cup \{Z \rightarrow A\}$与$F$等价。

如果$G^+=F^+$，就称F与G等价。 $G^+=F^+$的充分必要条件是$F \subseteq G^+$和$G \subseteq F^+$

‹#›

举例

- S(SNO, SNAME, SEDPT, MN, CNAME, G)
- F={SNO→SDEPT, SNO → SNAME, SDEPT → MN, SNO → MN, (SNO,SNAME) → SDEPT, (SNO,CNAME) → G}

① 求最小函数依赖集合和所有键

F_{min}={SNO → SDEPT, SNO → SNAME, SDEPT → MN, (SNO,CNAME)→ G}

键key为(SNO, CNAME)

② 函数依赖左部相同分组

{SNO → SDEPT, SNO → SNAME}　S1(SNO, SDEPT, SNAME)

{ SDEPT → MN}　S2(SDEPT, MN)

{(SNO,CNAME) → G}　S3(SNO, CNAME, G)

③ 不存在子模式的包含情况；　④ S3中含有key，分解终止。

‹#›

雨课堂 Rain Classroom

判断无损连接分解的算法

- 已知R<U,F>， U={A,B,C,D,E}
- F={AB →C, C →D, D →E}
- R的一个分解R1(A,B,C),R2(C,D),R3(D,E)

	A	B	C	D	E
R1(A,B,C)	a_1	a_2	a_3	~~b_{14}~~ a_4	~~b_{15}~~ a_5
R_2(C,D)	b_{21}	b_{22}	a_3	a_4	~~b_{25}~~ a_5
R3(D,E)	b_{31}	b_{32}	b_{33}	a_4	a_5

- 已知R<U,F>， U={A,B,C,D,E}
- F={A→C,B→C, C →D, DE →C,CE →A}
- R的一个分解

  R1(A,D),R2(A,B),R3(B,E),R4(C,D,E),R5(A,E)

Review

- Third Normal Form (3NF)
- 任何关系模式都可以保持函数依赖和无损连接分解到3NF
- 保持函数依赖和无损连接的3NF模式分解算法
- 判断无损连接分解的算法

‹#›

单选题 1分

互动交流一

关系模式R(ABCD), F={B→D,AB→C}，R最高属于第几范式？

A R属于3NF

B R属于BCNF

C R属于1NF

D 以上答案都不正确

提交

‹#›

单选题 1分

互动交流二

关系模式R(ABCD), F={B→D,D→B,AB→C}，R最高属于第几范式?

A R属于3NF

B R属于BCNF

C R属于1NF

D 以上答案都不正确

提交

单选题 1分

互动交流三

关系模式R(ABC),F={A→B,B→A,A→C}，R最高属于第几范式?

A R属于3NF

B R属于BCNF

C R属于1NF

D 以上答案都不正确

提交

单选题 1分

互动交流四

关系模式R(ABCD), F={B→D,AB→C}，在保证
的前提下，将它分解为一组3NF关系模式，

A. R1(A,B), R2(B,C), R2(B,D)

B. R1(A,B), R2(B,C), R2(C,D)

C. R1(A,B,D), R2(C,D)

D. R1(A,B,C), R2(B,D)

提交

‹#›

单选题 1分

互动交流五

关系模式R(ABCDE), F={B→D,AB→C}，在保
的前提下，将它分解为一组3NF关系模式，

A. R1(A,B,C), R2(B,D)

B. R1(A,B,C), R2(B,D), R3(B,E)

C. R1(A,B,C), R2(B,D), R3(A,B,E)

D. R1(A,B,C), R2(B,D), R3(A,E)

提交

‹#›

主观题 10分

互动交流六

某学校设计了关系模式:

学生（身份证号，学号，姓名，籍贯，家庭住址）

这个关系模式是BCNF的吗？它是否存在插入、删除和更新异常等问题？

提交

‹#›

主观题 10分

互动交流七

某学校设计了关系模式:

学生（学号、课号、课程名称、成绩）

这个关系模式是BCNF的吗？如果不是，可否在保证无损连接和函数依赖的前提下，将它分解为一组BCNF的关系模式。

提交

‹#›

Week11_Handout

Database System-Schema+Design part7

‹#›

Schema Design and Refinement

- How do we obtain a good design?
- Functional dependencies & keys
- Desirable properties of schema refinement
- Boyce Codd Normal Form (BCNF)
- Third Normal Form (3NF) and 3NF Decomposition
- Fourth Normal Form (4NF)

‹#›

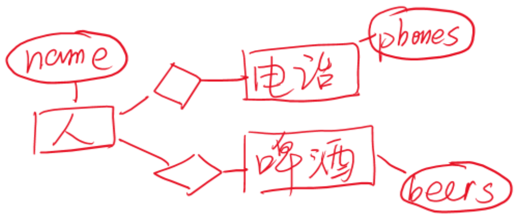
Multivalued Dependencies

Fourth Normal Form

‹#›

A New Form of Redundancy

- Multivalued dependencies (MVD’s) express a condition among tuples of a relation that exists when the relation is trying to represent more than one many-many relationship.
- Certain attributes are independent of one another, and their values must appear in all combinations.



name	phones	beers
张三	{p1, p2}	{b1, b2}
李四	{p1, p3}	{b2, b4, b5}

‹#›

Example-Tuples Implied by Independence

Drinkers(name, addr, phones, beersLiked)

- A drinker's phones are independent of the beers they like.
- Thus, each of a drinker's phones appears with each of the beers they like in all combinations.
- This repetition is unlike FD redundancy.
  - Name → addr is the only FD.

If we have tuples:

Name	addr	phones	beersLiked
sue	a	p1	b1
sue	a	p2	b2
sue	a	p2	b1
sue	a	p1	b2

Then these tuples must also be in the relation.

‹#›

Definition of MVD

- A *multivalued dependency* (MVD) $X \twoheadrightarrow Y$ is an assertion that if two tuples of a relation agree on all the attributes of *X*, then their components in the set of attributes *Y* may be swapped, and the result will be two tuples that are also in the relation.

‹#›

Example

Drinkers(name, addr, phones, beersLiked)

name	addr	phones	beersLiked
sue	a	p1	b1
sue	a	p2	b2
sue	a	p2	b1
sue	a	p1	b2

- MVD

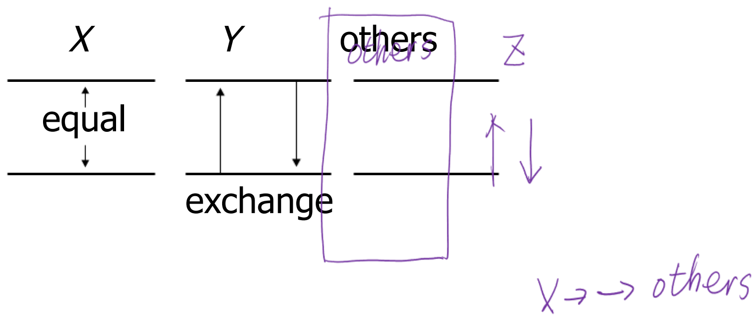
  name $\rightarrow\rightarrow$ phones

  and the MVD

  name $\rightarrow\rightarrow$ beersLiked.

‹#›

Picture of MVD $X \rightarrow\rightarrow Y$



‹#›

MVD Rules

- Every FD is an MVD (*promotion*).
  - If $X \to Y$, then swapping Y's between two tuples that agree on X doesn't change the tuples.
  - Therefore, the "new" tuples are surely in the relation, and we know $X \to\to Y$.
- *Complementation* : If $X \to\to Y$, and Z is all the other attributes, then $X \to\to Z$.

‹#›

Splitting Doesn't Hold

- Like FD's, we cannot generally split the left side of an MVD.
- But unlike FD's, we cannot split the right side either --- sometimes you have to leave several attributes on the the right side.

If $AB \to C$ then $A \to C$, $B \to C$ 错误	If $AB \to\to C$ then $A \to\to C$, $B \to\to C$ 错误
If $A \to BC$ then $A \to B$, $A \to C$ 正确	If $A \to\to BC$ then $A \to\to B$, $A \to\to C$ 错误

‹#›

Example

Drinkers(name, areaCode, phone, beersLiked, manf)

- A drinker can have several phones, with the number divided between areaCode and phone (last 7 digits).
- A drinker can like several beers, each with its own manufacturer.

‹#›

Example, Continued

- Since the areaCode-phone combinations for a drinker are independent of the beersLiked-manf combinations, we expect that the following MVD's hold:

name →→ areaCode phone

name →→ beersLiked manf

‹#›

Example Data

Here is possible data satisfying these MVD’ s:

name	areaCode	phone	beersLiked	manf
Sue	650	555-1111	Bud	A.B.
Sue	650	555-1111	WickedAle	Pete’ s
Sue	415	555-9999	Bud	A.B.
Sue	415	555-9999	WickedAle	Pete’ s

But we cannot swap area codes or phones by themselves.
That is, neither name →→ areaCode nor name →→ phone holds for this relation.

Fourth Normal Form

- The redundancy that comes from MVD’ s is not removable by putting the database schema in BCNF.
- There is a stronger normal form, called 4NF, that (intuitively) treats MVD’ s as FD’ s when it comes to decomposition, but not when determining keys of the relation.

4NF Definition

- A relation *R* is in *4NF* if whenever $X \twoheadrightarrow Y$ is a nontrivial MVD, then *X* is a superkey.
  - *Nontrivial* means that:
    1. Y is not a subset of X, and
    2. X and Y are not, together, all the attributes.
  - Note that the definition of "superkey" still depends on FD's only.

Example

Drinkers(<u>name</u>, addr, <u>phones</u>, <u>beersLiked</u>)

FD: name -> addr

MVD's: name $\twoheadrightarrow$ phones

name $\twoheadrightarrow$ beersLiked

- Key is {name, phones, beersLiked}.
- All dependencies violate 4NF.

Drinkers & 4NF

Drinkers & BCNF

Example, Continued

- Decompose using name → addr:

1. Drinkers1(name, addr) ∈4NF
   - In 4NF; only dependency is ~~name → addr~~.
2. Drinkers2(name, phones, beersLiked) ∉4NF
   - Not in 4NF. MVD's ~~name →→ phones~~ and ~~name →→ beersLiked~~ apply. No FD's, so all three attributes form the key.

Example: Decompose Drinkers2

- Either MVD name →→ phones or name →→ beersLiked tells us to decompose to:
  - Drinkers3(name, phones) ∈4NF
  - Drinkers4(name, beersLiked) ∈ 4NF

Decomposition and 4NF

- If $X \rightarrow\rightarrow Y$ is a 4NF violation for relation R, we can decompose R using the same technique as for BCNF.

1. XY is one of the decomposed relations.
2. All but $Y - X$ is the other.

‹#›

BCNF Versus 4NF

- Remember that every FD $X \rightarrow Y$ is also an MVD, $X \rightarrow\rightarrow Y$.
- Thus, if R is in 4NF, it is certainly in BCNF.
  - Because any BCNF violation is a 4NF violation.
- But R could be in BCNF and not 4NF, because MVD’s are “invisible” to BCNF.

‹#›

本章主要知识点

- Functional dependencies
  - 函数依赖：属性之间的约束，应用背景领域信息，通过用户得到
- Keys
  - 键或候选码：可以通过函数依赖求出。一个关系模式可以有多个键，在工程设计时一个关系模式只能指定一个键为主键，主键不能为空值，也不能出现重复值（实体完整性）
- Various normal forms
  - BCNF, 3rd normal form, 4th normal form. 关系模式如果不能达到3NF以上，会出现冗余和各种异常（Anomalies）
- schema refinement
  - Lossless Decompositions 无损分解，“lossless” joins 无损连接
  - Dependency Preserving 保持函数依赖

‹#›

单选题 1分

互动交流一

Relation R(A，B，C) satisfies the multi-valued dependency A→→B, and has (possibly among others) the following tuples in its current instance: (0, 1, 2), (0, 3, 4), and (1, 5, 6). Which of the following tuples is not necessarily in the current instance of R?

A. (0; 1; 4)

B. (0; 3; 2)

C. (0; 5; 2)

D. None of the above

提交

‹#›

多选题 1分

互动交流二

一个关系模式：学生（学号，社团，课程）存在多值函数依赖？

A 存在社团对学号的多值函数依赖

B 存在社团对课程的多值函数依赖

C 存在课程对学号的多值函数依赖

D 存在学号对课程的多值函数依赖

提交

‹#›

单选题 1分

互动交流三

一个关系模式：学生（学号，社团，课程）它是否为4NF?

A 它是BCNF？它也是4NF?

B 它是BCNF？它不是4NF?

C 它不是BCNF？它也不是4NF?

D 它不是BCNF？但它是4NF?

提交

‹#›

主观题 10分

互动交流四

一个关系模式：学生（学号，社团，课程），如果这个关系模式存在多值函数依赖，请对模式求精，去除这种多值函数依赖。求精后的模式由哪些关系模式组成？

提交

‹#›