数据库系统概论 An Introduction to Database System

第六章 关系数据理论

第六章 关系数据理论

- 6.1 问题的提出
- 6.2 规范化
- 6.3 数据依赖的公理系统
- *6.4 模式的分解
- 6.5 小结

6.1 问题的提出

关系数据库逻辑设计

■ 数据库逻辑设计的工具——关系数据库的规范化理论

问题的提出

- 一、概念回顾
- 二、关系模式的形式化定义
- 三、什么是数据依赖
- 四、关系模式的简化定义
- 五、数据依赖对关系模式影响

二、关系模式的形式化定义

关系模式由五部分组成,即它是一个五元组:

R(U, D, DOM, F)

R: 关系名

U: 组成该关系的属性名集合

D: 属性组U中属性所来自的域

DOM: 属性向域的映象集合

F: 属性U上的一组数据依赖

三、什么是数据依赖

- ❖数据依赖:属性与属性之间的约束关系。它是数据库模式设计的关键
- ❖通过属性间值的相等与否体现数据间的相互联系
- *它是现实世界属性间相互联系的抽象
- ❖它是数据内在的性质
- ❖它是语义的体现

什么是数据依赖(续)

数据依赖的类型

- ❖ 函数依赖(Functional Dependency,简记为FD)
- ❖ 多值依赖(Multivalued Dependency,简记为MVD)
- * 其他

四、关系模式的简化表示

* 关系模式R(U, D, DOM, F) 简化为一个三元组:

R(U,F)

❖ 当且仅当U上的一个关系「满足F时,「称为关系模式 R (U, F) 的一个关系

五、数据依赖对关系模式的影响

[例1]建立一个描述学校教务的数据库: 学生的学号(Sno)、所在系(Sdept)、 系所在地(Sloc)、课程号(Cno)、 成绩(Grade)

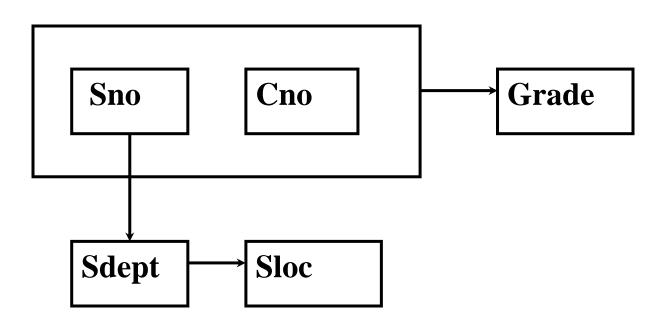
单一的关系模式: Student <U、F>

U = { Sno, Sdept, Sloc, Cno, Grade }

数据依赖对关系模式的影响 (续)

属性组U上的一组函数依赖F:

```
F = { Sno → Sdept, Sdept → Mname, (Sno, Cname) → Grade }
```



关系模式Student<U,F>中存在的问题

- 1. 数据冗余太大
- 2. 更新异常 (Update Anomalies)
- 3. 插入异常 (Insertion Anomalies)
- 4. 删除异常 (Deletion Anomalies)

数据依赖对关系模式的影响(续)

结论:

- Student关系模式不是一个好的模式。
- "好"的模式:

不会发生插入异常、删除异常、更新异常,数据冗余应尽可能少

原因:由存在于模式中的某些数据依赖引起的

解决方法: 通过分解关系模式来消除其中不合适的数据依赖

分解关系模式

❖把这个单一模式分成3个关系模式:

```
S (Sno, Sdept, Sno \rightarrow Sdept);
SC (Sno, Cno, Grade, (Sno, Cno) \rightarrow Grade);
DEPT (Sdept, Sloc, Sdept\rightarrow Sloc)
```

第六章 关系数据理论

- 6.1 问题的提出
- 6.2 规范化
- 6.3 数据依赖的公理系统
- *6.4 模式的分解
- 6.5 小结

6.2 规范化

规范化理论正是用来改造关系模式,通过分解关系模式来消除其中不合适的数据依赖,以解决插入异常、删除异常、更新异常和数据冗余问题。

6.2 规范化

- 6.2.1 函数依赖
- 6.2.2 码
- 6.2.3 范式
- 6.2.4 2NF
- 6.2.5 3NF
- 6.2.6 BCNF
- 6.2.7 多值依赖
- 6.2.8 4NF
- 6.2.9 规范化小结

6.2.1 函数依赖

- ❖函数依赖
- ❖平凡函数依赖与非平凡函数依赖
- ◆完全函数依赖与部分函数依赖
- *传递函数依赖

一、函数依赖

定义6.1 设R(U)是一个属性集U上的关系模式,X和Y是U的子集。

若对于R(U)的任意一个可能的关系r,r中不可能存在两个元组在X上的属性值相等,而在Y上的属性值不等,则称"X函数

确定Y"或 "Y函数依赖于X",记作X→Y。

二、平凡函数依赖与非平凡函数依赖

在关系模式R(U)中,对于U的子集X和Y,如果X \rightarrow Y,但Y \subseteq X,则称X \rightarrow Y是非平凡的函数依赖若X \rightarrow Y,但Y \subseteq X,则称X \rightarrow Y是平凡的函数依赖

❖ 例: 在关系SC(Sno, Cno, Grade)中,

非平凡函数依赖: (Sno, Cno) → Grade

平凡函数依赖: (Sno, Cno) → Sno

 $(Sno, Cno) \rightarrow Cno$

平凡函数依赖与非平凡函数依赖(续)

- 若 $X \rightarrow Y$,则X称为这个函数依赖的决定属性组,也称为决定因素(Determinant)。
- 若 $X \rightarrow Y$, $Y \rightarrow X$,则记作 $X \leftarrow \rightarrow Y$ 。
- 若Y不函数依赖于X,则记作X→Y。

三、完全函数依赖与部分函数依赖

定义6.2 在R(U)中,如果X→Y,并且对于X的任何一个真子集X',都有X'→、Y,则称Y对X完全函数依赖,记作 X $\stackrel{\mathsf{F}}{\longrightarrow}$ Y。 若X→Y,但Y不完全函数依赖于X,则称Y对X部分函数依赖,记作X $\stackrel{\mathsf{P}}{\longrightarrow}$ Y。

完全函数依赖与部分函数依赖(续)

[例1] 中(Sno,Cno)^F→Grade是完全函数依赖, (Sno,Cno)^P→Sdept是部分函数依赖

因为Sno →Sdept成立,且Sno是(Sno, Cno)的真子集

四、传递函数依赖

定义**6.3** 在R(U)中,如果X \rightarrow Y,(Y \searrow X),Y \rightarrow X Y \rightarrow Z,则称Z对X传递函数依赖。

记为: X ^{传递} Z

注: 如果 $Y \rightarrow X$,即 $X \leftarrow \rightarrow Y$,则Z直接依赖于X。

例: 在关系Std(Sno, Sdept, Sloc)中,有: Sno → Sdept, Sdept → Sloc Sloc传递函数依赖于Sno

6.2 规范化

- 6.2.1 函数依赖
- 6.2.2 码
- 6.2.3 范式
- 6.2.4 2NF
- 6.2.5 3NF
- 6.2.6 BCNF
- 6.2.7 多值依赖
- 6.2.8 4NF
- 6.2.9 规范化小结

6.2.2 码 (使用函数依赖的概念来定义)

定义6.4 设K为R<U,F>中的属性或属性组合。若 $K \xrightarrow{F} U$,

则K称为R的侯选码(Candidate Key)。

若候选码多于一个,则选定其中的一个做为主码(Primary Key)

码(续)

- * 主属性与非主属性
 - 包含在任何一个候选码中的属性 , 称为主属性 (Prime attribute)
 - 不包含在任何码中的属性称为非主属性(Nonprime attribute)
 或非码属性(Non-key attribute)
- * 全码
 - 整个属性组是码,称为全码(All-key)

码(续)

[例2]

关系模式S(<u>Sno</u>,Sdept,Sage),单个属性Sno是码, SC(<u>Sno,Cno</u>,Grade)中,(Sno,Cno)是码 [例3]

关系模式R(P,W,A)

P: 演奏者 W: 作品 A: 听众

一个演奏者可以演奏多个作品

某一作品可被多个演奏者演奏

听众可以欣赏不同演奏者的不同作品

码为(P, W, A), 即All-Key

外部码

定义6.5 关系模式 R 中属性或属性组X 并非 R的码,但 X 是另一个关系模式的码,则称 X 是R 的外部码 (Foreign key) 也称外码

- ❖ 如在SC (Sno, Cno, Grade) 中, Sno不是码, 但 Sno是关系模式S (Sno, Sdept, Sage) 的码,则 Sno是关系模式SC的外部码
- * 主码与外部码一起提供了表示关系间联系的手段

6.2 规范化

- 6.2.1 函数依赖
- 6.2.2 码
- 6.2.3 范式
- 6.2.4 2NF
- 6.2.5 3NF
- 6.2.6 BCNF
- 6.2.7 多值依赖
- 6.2.8 4NF
- 6.2.9 规范化小结

6.2.3 范式

- * 范式是符合某一种级别的关系模式的集合
- *满足不同程度要求的为不同范式
- ❖ 范式的种类:

第一范式(1NF)

第二范式(2NF)

第三范式(3NF)

BC范式(BCNF)

第四范式(4NF)

第五范式(5NF)

6.2.3 范式

* 各种范式之间存在联系:

 $1NF \supset 2NF \supset 3NF \supset BCNF \supset 4NF \supset 5NF$

- ※某一关系模式R为第n范式,可简记为R∈nNF。
- ❖ 一个低一级范式的关系模式,通过模式分解可以转换为若 干个高一级范式的关系模式的集合,这种过程就叫规范化

6.2 规范化

- 6.2.1 函数依赖
- 6.2.2 码
- 6.2.3 范式
- 6.2.4 2NF
- 6.2.5 3NF
- 6.2.6 BCNF
- 6.2.7 多值依赖
- 6.2.8 4NF
- 6.2.9 规范化小结

6.2.4 2NF

❖ 1NF的定义

如果一个关系模式R的所有属性都是不可分的基本数据项,则R∈1NF

SUPERVISOR	SPECIALITY	POSTGRADUATE		
		PG1	PG2	
张清玫	信息专业	李勇	刘晨	
刘逸	信息专业	王敏		小表

- ※第一范式是对关系模式的最起码的要求。不满足第一范式的数据库模式不能称为关系数据库
- * 但满足第一范式的关系模式并不一定是一个好的关系模式

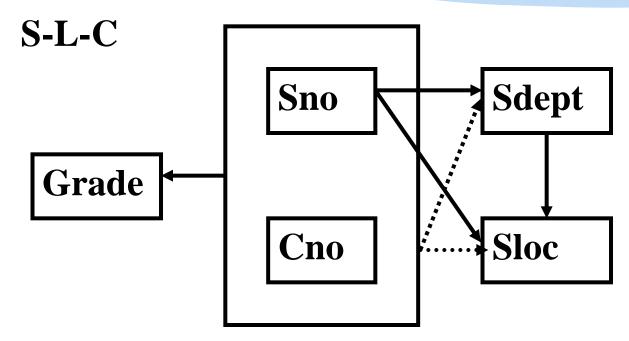
2NF(续)

[例4] 关系模式 S-L-C(Sno, Sdept, Sloc, Cno, Grade) Sloc为学生住处,假设每个系的学生住在同一个地方

* 函数依赖包括:

(Sno, Cno) \vdash Grade Sno → Sdept (Sno, Cno) $\stackrel{P}{\longrightarrow}$ Sdept Sno → Sloc (Sno, Cno) $\stackrel{P}{\longrightarrow}$ Sloc Sdept → Sloc

2NF(续)



- ❖ S-L-C的码为(Sno, Cno)
- ❖ S-L-C满足第一范式。
- ❖ 非主属性Sdept和Sloc部分函数依赖于码(Sno, Cno)

S-L-C不是一个好的关系模式(续)

- (1) 数据冗余度大
- (2) 插入异常
- (3) 删除异常
- (4) 修改复杂

S-L-C不是一个好的关系模式(续)

❖ 原因

Sdept、 Sloc部分函数依赖于码。

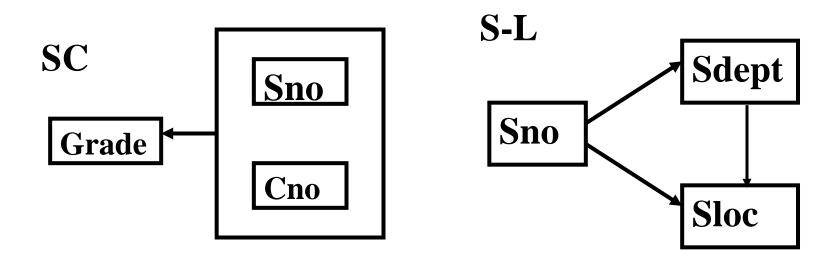
*解决方法

S-L-C分解为两个关系模式,以消除这些部分函数依赖

SC (Sno, Cno, Grade)

S-L (Sno, Sdept, Sloc)

函数依赖图:



- ❖关系模式SC的码为 (Sno, Cno)
- ❖关系模式S-L的码为Sno
- *这样非主属性对码都是完全函数依赖

❖2NF的定义

定义6.6 若R∈1NF,且每一个非主属性完全函数依赖于码,则R∈2NF。

例: S-L-C(Sno, Sdept, Sloc, Cno, Grade) ∈1NF S-L-C(Sno, Sdept, Sloc, Cno, Grade) ∈2NF SC (Sno, Cno, Grade) ∈ 2NF S-L (Sno, Sdept, Sloc) ∈ 2NF

* 采用投影分解法将一个1NF的关系分解为多个2NF的关系,可以在一定程度上减轻原1NF关系中存在的插入异常、删除异常、数据冗余度大、修改复杂等问题。

❖ 将一个1NF关系分解为多个2NF的关系,并不能完全消除 关系模式中的各种异常情况和数据冗余。

6.2 规范化

- 6.2.1 函数依赖
- 6.2.2 码
- 6.2.3 范式
- 6.2.4 2NF
- 6.2.5 3NF
- 6.2.6 BCNF
- 6.2.7 多值依赖
- 6.2.8 4NF
- 6.2.9 规范化小结

6.2.5 3NF

❖3NF的定义

定义6.7 关系模式R < U,F > 中若不存在这样的码X、属性组Y及非主属性Z($Z \succeq Y$),使得 $X \rightarrow Y$, $Y \rightarrow Z$ 成立, $Y \rightarrow X$,则称R < U, $F > \in 3NF$ 。

■若**R**∈3NF,则每一个非主属性既不部分依赖于码也不 传递依赖于码。

例: 2NF关系模式S-L(Sno, Sdept, Sloc)中

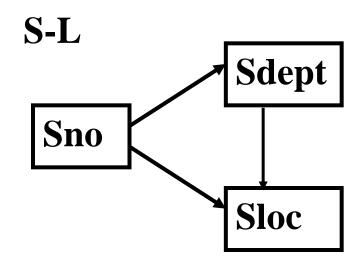
■ 函数依赖:

Sno→Sdept

Sdept → Sno

Sdept→Sloc

可得:



Sno^{传递} Sloc,即S-L中存在非主属性对码的传递函数依赖,S-L *₹ 3NF*

❖解决方法

采用投影分解法,把S-L分解为两个关系模式,以消除传递函数依赖:

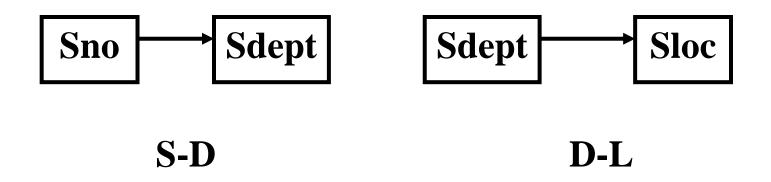
S-D (Sno, Sdept)

D-L (Sdept, Sloc)

S-D的码为Sno, D-L的码为Sdept。

■ 分解后的关系模式S-D与D-L中不再存在传递依赖

S-D的码为Sno, D-L的码为Sdept



S-L(Sno, Sdept, Sloc) ∈ 2NF
 S-L(Sno, Sdept, Sloc) ∈ 3NF
 S-D(Sno, Sdept) ∈ 3NF
 D-L(Sdept, Sloc) ∈ 3NF

❖ 采用投影分解法将一个2NF的关系分解为多个3NF的关系,可以在一定程度上解决原2NF关系中存在的插入异常、删除异常、数据冗余度大、修改复杂等问题。

❖ 将一个2NF关系分解为多个3NF的关系后,仍然不能完全消除 关系模式中的各种异常情况和数据冗余。

6.2 规范化

- 6.2.1 函数依赖
- 6.2.2 码
- 6.2.3 范式
- 6.2.4 2NF
- 6.2.5 3NF
- 6.2.6 BCNF
- 6.2.7 多值依赖
- 6.2.8 4NF
- 6.2.9 规范化小结

6.2.6 BC范式 (BCNF)

*定义6.8 关系模式R<U, F>∈1NF, 若X→Y且Y ⊆ X时X必含有码,则R<U, F>∈BCNF。

❖等价于:每一个决定属性因素都包含码

*****若R∈BCNF

- 所有非主属性对每一个码都是完全函数依赖
- 所有的主属性对每一个不包含它的码,也是完全函数依赖
- 没有非主属性对码是传递函数依赖
- 所有的主属性对每一个不包含它的码,不存在传递函数依赖
- 没有任何属性完全函数依赖于非码的任何一组属性

♦ R ∈ BCNF
$$\xrightarrow{\hat{R}\hat{Y}}$$
 R ∈ 3NF

[例5] 关系模式C(Cno,Cname,Pcno)

- C ∈ 3NF
- $C \in BCNF$

[例6] 关系模式S (Sno, Sname, Sdept, Sage)

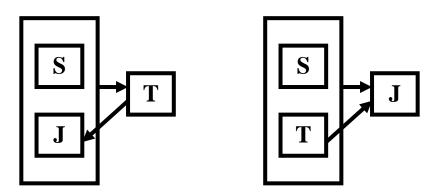
- 假定S有两个码Sno, Sname
- **S**∈3NF。
- \blacksquare S \in BCNF

[例7] 关系模式SJP(S, J, P)

- ■S是学生,J是课程,P是名次
- ■函数依赖: (S, J) →P; (J, P) →S
- (S, J) 与 (J, P) 都可以作为候选码,属性相交
- \blacksquare SJP \in 3NF,
- ■SJP ∈ BCNF

[例8]关系模式STJ(S,T,J),S表示学生,T表示教师,J表示课程。

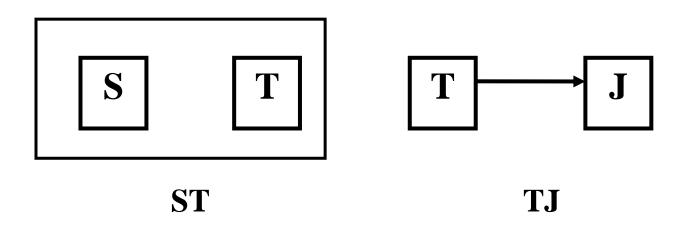
- 函数依赖: (S, J)→T, (S, T)→J, T→J(假设每个教师教一门课)
- (S, J)和(S, T)都是候选码



STJ中的函数依赖

- STJ ∈ 3NF,没有任何非主属性对码传递依赖或部分依赖
- STJ∈BCNF, T是决定因素, T不包含码

解决方法:将STJ分解为二个关系模式: ST(S, T) ∈ BCNF, TJ(T, J) ∈ BCNF



没有任何属性对码的部分函数依赖和传递函数依赖

3NF与BCNF的关系

$$R$$
 ∈ BCNF $\xrightarrow{\text{$\hat{\Lambda}$}}$ R ∈ 3NF

❖如果R∈3NF,且R只有一个候选码

R ∈ BCNF
$$\xrightarrow{\hat{\Sigma}\hat{Y}}$$
 R ∈ 3NF

6.2 规范化

- 6.2.1 函数依赖
- 6.2.2 码
- 6.2.3 范式
- 6.2.4 2NF
- 6.2.5 3NF
- 6.2.6 BCNF
- 6.2.7 多值依赖
- 6.2.8 4NF
- 6.2.9 规范化小结

6.2.7 多值依赖

[例9] 学校中某一门课程由多个教师讲授,他们使用相同的一套参考书。每个教员可以讲授多门课程,每种参考书可以供多门课程使用。

课程C	教员T	参考书B
物理	李勇王军	普通物理学 光学原理 物理习题集
数学 	{ 李 勇 } { 张 平}	数学分析 微分方程 高等代数
计算数学	{张平} (周峰)	{数学分析
;	·	;

课程C	教员T	参考书B
物理	李 勇	普通物理学
物 理	李 勇	光学原理
物 理	李 勇	物理习题集
物 理	王军	普通物理学
物 理	王军	光学原理
物 理	王军	物理习题集
数 学	李 勇	数学分析
数 学	李 勇	微分方程
数 学	李 勇	高等代数
数 学	张 平	数学分析
数 学	张 平	微分方程
数 学	张 平	高等代数

- ❖ Teaching ∈ BCNF
- ❖ Teaching具有唯一候选码(C, T, B), 即全码

Teaching模式中存在的问题

- (1)数据冗余度大
- (2)插入操作复杂
- (3) 删除操作复杂
- (4) 修改操作复杂



* 定义6.9

设R(U)是一个属性集U上的一个关系模式, X、 Y和Z是U的子集,并且Z=U-X-Y。关系模式R(U)中多值依赖 X→→Y成立,当且仅当对R(U)的任一关系r,给定的一对(x,z)值,有一组Y的值,这组值仅仅决定于x值而与z值无关

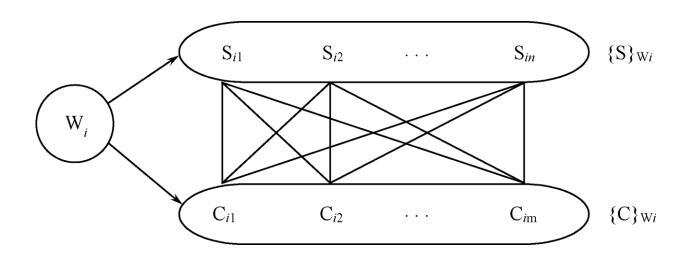
❖ 例 Teaching (C, T, B)

[例10] 关系模式WSC(W,S,C)

- W表示仓库,S表示保管员,C表示商品
- 假设每个仓库有若干个保管员,有若干种商品
- 每个保管员保管所在的仓库的所有商品
- 每种商品被所有保管员保管

W	S	C
W1	S1	C1
W1	S 1	C2
W1	S1	C3
W1	S2	C1
W1	S2	C2
W1	S2	C3
W2	S3	C4
W2	S3	C5
W2	S4	C4
W2	S4	C5

用下图表示这种对应,形成了一个完全二分图



$$W \rightarrow \rightarrow S \sqcap W \rightarrow \rightarrow C$$

多值依赖的性质

- (1) 多值依赖具有对称性
 - 若X→→Y,则X→→Z,其中Z=U-X-Y

- (4) 若 $X \rightarrow Y$, $X \rightarrow Z$, 则 $X \rightarrow Y \cap Z$ 。
- (5) 若 $X \rightarrow Y$, $X \rightarrow Z$, 则 $X \rightarrow Y Z$, $X \rightarrow Z Y$ 。

6.2 规范化

- 6.2.1 函数依赖
- 6.2.2 码
- 6.2.3 范式
- 6.2.4 2NF
- 6.2.5 3NF
- 6.2.6 BCNF
- 6.2.7 多值依赖
- 6.2.8 4NF
- 6.2.9 规范化小结

6.2 规范化

- 6.2.1 函数依赖
- 6.2.2 码
- 6.2.3 范式
- 6.2.4 2NF
- 6.2.5 3NF
- 6.2.6 BCNF
- 6.2.7 多值依赖
- 6.2.8 4NF
- 6.2.9 规范化小结

6.2.9 规范化小结

* 关系数据库的规范化理论是数据库逻辑设计的工具

❖目的:尽量消除插入、删除异常,修改复杂,数据冗余

*基本思想:逐步消除数据依赖中不合适的部分

■ 实质: 概念的单一化

规范化小结(续)

* 关系模式规范化的基本步骤

消除决定属性 集非码的非平 凡函数依赖

1NF

- ↓消除非主属性对码的部分函数依赖
- 2NF
- →消除非主属性对码的传递函数依赖
- 3NF
 - ↓消除主属性对码的部分和传递函数依赖

BCNF

- ↓消除非平凡且非函数依赖的多值依赖
- 4NF

规范化小结(续)

- * 不能说规范化程度越高的关系模式就越好
- ❖ 在设计数据库模式结构时,必须对现实世界的实际情况和用户应用需求作进一步分析,确定一个合适的、能够反映现实世界的模式
- *上面的规范化步骤可以在其中任何一步终止

第六章 关系数据理论

- 6.1 问题的提出
- 6.2 规范化
- 6.3 数据依赖的公理系统
- *6.4 模式的分解
- 6.5 小结

第六章 关系数据理论

- 6.1 问题的提出
- 6.2 规范化
- 6.3 数据依赖的公理系统
- *6.4 模式的分解
- 6.5 小结

下课了。。。

休息一会儿。。。