

数据库原理

Theory of Database

李静

信息科学与技术学院

问题的提出

S-L-C(Sno,Sname,Ssex,Sdept,SLoc,Cno,Grade)

主码(sno,cno)

Sno	Sname	Ssex	Sdept	SLoc	Cno	Grade
9512101	李勇	男	计算机系	2公寓	c01	90
9512101	李勇	男	计算机系	2公寓	c02	86
9512101	李勇	男	计算机系	2公寓	c06	NULL
9512102	刘晨	男	计算机系	2公寓	c02	78
9512102	刘晨	男	计算机系	2公寓	c04	66
9521102	吴宾	女	信息系	1公寓	c01	82
9521102	吴宾	女	信息系	1公寓	c02	75
9521102	吴宾	女	信息系	1公寓	c04	92
9521102	吴宾	女	信息系	1公寓	c05	50
9521103	张海	男	信息系	1公寓	c02	68
9521103	张海	男	信息系	1公寓	c06	NULL
9531101	钱小平	女	数学系	1公寓	c01	80
9531101	钱小平	女	数学系	1公寓	c05	95

问题的提出

❖ 该关系实例可能出现：

- 学生的基本信息和宿舍楼被重复存储多次；
- 当修改某学生的系名时，还要修改宿舍楼的信息；
- 如果某学生没有选修课程，或某门课程未被任何学生选修时，则该学生或该课程信息不能存入数据库，因为主码值不能为空；
- 当一学生的选修课程信息都被删除时，则该学生的信息可能会丢失。

冗余存储

更新异常

插入异常

删除异常

问题的提出

❖ **数据冗余**是指同一信息在数据库中存储了多个副本。

它可能引起下列问题：

- **冗余存储**：信息被重复存储，导致浪费大量存储空间。
- **更新异常**：当重复信息的一个副本被修改，所有副本都必须进行同样的修改。因此当更新数据时，系统要付出很大的代价来维护数据库的完整性，否则会面临数据不一致的危险。
- **插入异常**：只有当一些信息事先已经存放在数据库中时，另外一些信息才能存入数据库中。
- **删除异常**：删除某些信息时可能丢失其它信息。

问题的提出

- ❖ 结论：此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，
- ❖ 原因：由存在于模式中的某些数据依赖引起
- ❖ 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。

本章要解决的两个问题

- ❖ 如何判断一个数据库模式是“好”的模式
- ❖ 如何设计出一个“好”模式？

第6章 关系数据库规范化理论

❖ 6.1 函数依赖

- 6.1.1 函数依赖基本概念
- 6.1.2 一些术语和符号
- 6.1.3 为什么要讨论函数依赖

❖ 6.2 关系规范化

- 6.2.1 关系模式中的码
- 6.2.2 范式



6.1 函数依赖

- ❖ 数据的语义不仅表现为完整性约束，对关系模式的设计也提出了一定的要求。
- ❖ 如何构造一个合适的关系模式，应构造几个关系模式，每个关系模式由哪些属性组成等，都是数据库设计问题，确切地讲是关系数据库的逻辑设计问题。

6.1.1 函数依赖基本概念

$$Y=f(X)$$

$$\text{省}=f(\text{城市})$$

只要给出一个具体的城市值，就会有唯一一个省值和它对应。

把 X 函数决定 Y ，或 Y 函数依赖于 X 表示为：

$$X \rightarrow Y$$

6.1.1 函数依赖基本概念

❖ 如果有关系模式 $R(A_1, A_2, \dots, A_n)$, X 和 Y 为 $\{A_1, A_2, \dots, A_n\}$ 的子集, 则对于关系 R 中的任意一个 X 值, 都只有一个 Y 值与之对应, 则称 X 函数决定 Y , 或 Y 函数依赖于 X 。

示例

例1：对学生关系模式

Student (Sno, Sname, Sdept, Sage)

有以下依赖关系：

$Sno \rightarrow Sname, Sno \rightarrow Sdept, Sno \rightarrow Sage$

例2：对于选课关系

SC (Sno, Cno, Grade)

$(Sno, Cno) \rightarrow Grade$

函数依赖定义

- ❖ 定义：设有关系模式 $R(A_1, A_2, \dots, A_n)$ ， X 和 Y 均为 $\{A_1, A_2, \dots, A_n\}$ 的子集， r 是 R 的任一具体关系， t_1 、 t_2 是 r 中的任意两个元组；
- ❖ 如果由 $t_1[X] = t_2[X]$ 可推导出 $t_1[Y] = t_2[Y]$ ，则称 **X 函数决定 Y** ，或 **Y 函数依赖于 X** ，记为 $X \rightarrow Y$ 。

6.1.2 一些术语和符号

1. 如果 $X \rightarrow Y$ ，但 Y 不包含于 X ，则称

$X \rightarrow Y$ 是非平凡的函数依赖。

2. 如果 $X \rightarrow Y$ ，但 Y 包含于 X ，则称

$X \rightarrow Y$ 是平凡的函数依赖。

若无特别声明，讨论的都是非平凡的函数依赖。

3. 如果 $X \rightarrow Y$ ，则称 X 为决定因子， Y 为被决定因子。

4. 如果 $X \rightarrow Y$ ，并且 $Y \rightarrow X$ ，则记作 $X \longleftrightarrow Y$ 。

术语和符号（续）

5. 如果 $X \rightarrow Y$ ，并且对于 X 的一个任意真子集 X' 都有 $X' \not\rightarrow Y$ ，则称 Y 完全函数依赖于 X ：

$$X \xrightarrow{f} Y$$

如果 $X \rightarrow Y$ 成立，并且对于 X 的某个真子集 X' 有 $X' \rightarrow Y$ 成立，则称 Y 部分函数依赖于 X ：

$$X \xrightarrow{p} Y$$

6. 如果 $X \rightarrow Y$ （非平凡函数依赖，且 $Y \not\rightarrow X$ ） $Y \rightarrow Z$ ，则称 Z 传递函数依赖于 X ，记作：

$$X \xrightarrow{\text{传递}} Z$$

示例

❖ 例1:

关系模式SC (Sno, Sname, Cno, Grade) ,
主码为 (Sno, Cno) , 则函数依赖关系有:

$Sno \rightarrow Sname$

姓名函数依赖于学号

$(Sno, Cno) \xrightarrow{p} Sname$

姓名部分函数依赖于学号和课程号

$(Sno, Cno) \xrightarrow{f} Grade$

成绩完全函数依赖于学号和课程号

示例

❖ 例2:

关系模式S (Sno, Sname, Dept, Dept_master)，其中各属性分别为：学号、姓名、所在系和系主任（假设一个系只有一个主任），主码为Sno，则函数依赖关系有：

$Sno \xrightarrow{f} Sname$

姓名完全函数依赖于学号

由于： $Sno \xrightarrow{f} Dept$

所在系完全函数依赖于学号

$Dept \xrightarrow{f} Dept_master$

系主任完全函数依赖于系


所以有： $Sno \xrightarrow{\text{传递}} Dept_master$

系主任传递函数依赖于学号

6.2 关系规范化

- ❖ 关系规范化是指将有“不良”函数依赖的关系模式转换为良好的关系模式的理论。
- ❖ 这里涉及到范式的概念，不同的范式表示关系模式遵守的不同的规则。

问题的提出



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: 属性间数据的依赖关系集合

问题的提出

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 r 满足 F 时， r 称为关系模式 $R(U, F)$ 的一个关系。

6.2.1 关系模式中的码

候选码：设 K 为 $R(U, F)$ 中的属性或属性组，若 $K \xrightarrow{F} U$ (K 完全决定 U)，则 K 为 R 的**候选码**。

主码：关系 $R(U, F)$ 中可能有多个候选码，则选其中一个作为主码。

全码：候选码为整个属性组。

包含在任一候选码中的属性称为**主属性**，

不包含在任一候选码中的属性称为**非主属性**。

示例

例1：学生（学号，姓名，性别，身份证号，年龄，所在系）

❖ 候选码：学号，身份证号。

❖ 主码：“学号”。

❖ 主属性：学号，身份证号。

❖ 非主属性：姓名，性别，年龄，所在系

示例

例2. 选课（学号，课程号，考试次数，成绩）

❖ 设一个学生对一门课程可以有多次考试，每一次考试有一个考试成绩。

❖ 候选码：

（学号，课程号，考试次数），也为主码。

❖ 主属性：学号，课程号，考试次数

❖ 非主属性：成绩。

示例

例3. 有关系模式：授课（教师号，课程号，学年）

❖ 候选码：（教师号，课程号，学年）。

❖ 主码：同候选码。

❖ 主属性：教师号，课程号，学年。

❖ 非主属性：没有

❖ 是全码表。

外码

外码： 关系之间建立关联的属性（组）。

定义： 关系模式R中属性（组）X是另一个关系模式S的候选码（通常为主码），则称X为R的**外码**（**Foreign key**）。

例如： 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。

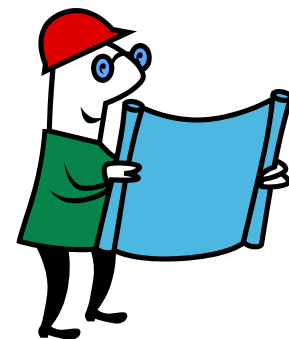
主码与外部码一起提供了表示关系间联系的手段。

6.2.2 范式

❖ 关系数据库中的关系要满足一定的要求，满足不同程度要求的为不同的**范式**

❖ **范式的种类：**

- 第一范式(1NF)
- 第二范式(2NF)
- 第三范式(3NF)
- 扩展的第三范式(BCNF)



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 $R \in nNF$ 。
- ❖ 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。

6.2.4 第一范式



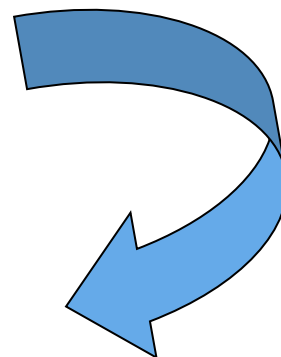
1NF的定义:

- 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 **$R \in 1NF$** 。
- 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。
- 但是满足第一范式的关系模式并不一定是一个好的关系模式。

第一范式

第一范式(1NF)：所有属性都不可分的关系

系名称	高级职称人数	
	教授	副教授
计算机系	6	10
信息管理学	3	5
电子与通讯系	4	8



系名称	教授人数	副教授人数
计算机系	6	10
信息管理学	3	5
电子与通讯系	4	8

第二范式



2NF的定义:

➡ 如果 $R(U, F) \in 1NF$, 并且 R 中的每个非主属性都完全函数依赖于码, 则 $R(U, F) \in 2NF$ 。

例: **S-L-C(Sno, Sdept, SLOC, Cno, Grade)**

因为有: $Sno \rightarrow SLOC$, 因此存在部分函数依赖, 该表不是2NF。

第二范式

- ❖ 若某个**1NF**的关系的主码只由一个码组成，那么这个关系就是**2NF**关系，但如果主码是由多个属性共同构成的复合主码，并且存在非主属性对主码的部分函数依赖，则这个关系就不是**2NF**。
- ❖ **第二范式的目标是：**将只部分依赖于主码（即依赖于主码的部分属性）的数据移到其他表中。

分解方法

- ❖ 首先，对于组成主码的属性集合的每一个子集，用它作为主码构成一个表。
- ❖ 然后，将依赖于这些主码的属性放置到相应的表中。
- ❖ 最后，去掉只由主码的子集构成的表。

分解S-L-C表

- ❖ 对**S-L-C** (**Sno**,**Sdept**,**SLOC**,**Cno**,**Grade**)
- ❖ 首先分解为如下形式的三张表：
 - **S-L** (**Sno**, ...)
 - **C** (**Cno**, ...)
 - **S-C** (**Sno**, **Cno**,...)

分解S-L-C表（续）

❖ 然后，将依赖于这些主码的属性放置到相应的表中，形成如下三张表：

- S-L (**Sno**, Sdept, Sloc)
- C (**Cno**)
- S-C (**Sno**, **Cno**, Grade)

❖ 最后，去掉只由主码的子集构成的表。

S-L-C (Sno, Sdept, SLOC, Cno, Grade)

分解S-L-C表（续）

❖ S-L-C关系模式最终分解的形式为：

- S-L (**Sno**, Sdept, Sloc)
- S-C (**Sno**, **Cno**, Grade)

分解后的函数依赖关系

❖ 分解后的关系模式的函数依赖关系：

❖ S-L (Sno, Sdept, Sloc) :

$Sno \rightarrow Sdept, Sno \rightarrow SLOC$ 是2NF

❖ S-C (Sno, Cno, Grade) :

$(Sno, Cno) \rightarrow Grade$ 是2NF

S-L (Sno, Sdept, SL0C) 存在问题

❖ 数据冗余：

有多少个学生就有多少个重复的Sdept和SL0C;

❖ 插入异常：

当新建一个系时，若还没有招收学生，则无法插入；

第三范式

第三范式 (3NF) : 如果 $R(U, F) \in 2NF$, 并且所有**非**主属性都不传递依赖于码, 则 $R(U, F) \in 3NF$ 。

S-L (Sno, Sdept, SL0C)

因为 $Sno \rightarrow Sdept$, $Sdept \rightarrow SL0C$,

所以 $Sno \xrightarrow{\text{传递}} SL0C$, 不是3NF。

第三范式

- ❖ 第三范式的目标是：去掉表中不依赖于主码的数据。
- ❖ 也就是说，在满足第二范式的实体中，非主属性不能依赖于另一个非主属性，即所有的非主属性应该直接依赖于全部的主属性（即必须完全依赖，这是2NF的要求），并且彼此之间无相互依赖关系（即不能存在部分依赖，这是3NF的要求）。

分解方法

- ❖ (1) 对于不是候选码的每个决定因子，从表中删去依赖于它的所有属性；
- ❖ (2) 新建一个表，新表中包含在原表中所有依赖于该决定因子的属性；
- ❖ (3) 将决定因子作为新表的主码。

S-L(Sno,Sdept,SLOC) Sno→Sdept Sdept→SLOC

S-D (Sno, Sdept)

S-L (Sdept, Sloc)

分解S-L关系模式

❖ S-L分解后的关系模式为：

S-D (Sno, Sdept)

S-L (Sdept, Sloc)

对 S-D, 有: $Sno \xrightarrow{f} Sdept$, 因此 S-D 是 3NF 的。

对 S-L, 有: $Sdept \xrightarrow{f} Sloc$, 因此 S-L 也是 3NF 的。

❖ S-L-C

(Sno, Sname, Ssex, Sdept, Sloc, Cno, Grade)

共分解为如下三个关系模式，每个关系模式都是3NF的。

❖ **S-D (Sno, Sname, Ssex, Sdept)**， Sdept为引用S-L关系模式的外码。

❖ **S-L (Sdept, Sloc)**， 没有外码。

❖ **S-C (Sno, Cno, Grade)**， Sno为引用S-D关系模式的外码。

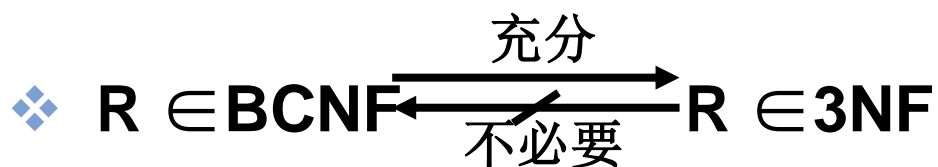
扩展的第三范式 (BCNF)

- ❖ 定义6.8 关系模式 $R\langle U, F \rangle \in 1NF$, 若 $X \twoheadrightarrow Y$ 且 $Y \subseteq X$ 时 X 必含有码, 则 $R\langle U, F \rangle \in BCNF$ 。
- ❖ 等价于: 非平凡函数依赖的每一个决定属性因素都包含候选码

BCNF (续)

❖ 若 $R \in \text{BCNF}$

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



BCNF (续)

[例5] 关系模式C (Cno, Cname, Pcnno)

- $C \in 3NF$
- $C \in BCNF$

[例6] 关系模式S (Sno, Sname, Sdept, Sage, fno)

- S有两个码Sno, fno(身份证号)
- $S \in 3NF$ 。
- $S \in BCNF$

BCNF（续）

[例7] 在关系模式**STJ**（**S**，**T**，**J**）中，**S**表示学生，**T**表示教师，**J**表示课程。每个教师只教一门课，一门课可以有多个老师，学生选定课程就固定了老师。

- 函数依赖：

$$(S, J) \rightarrow T, (S, T) \rightarrow J, T \rightarrow J$$

- 候选码： (S, J) 和 (S, T)

❖ **STJ** ∈ 3NF

- 没有任何非主属性对码传递依赖或部分依赖

❖ **STJ** 不满足 BCNF

- **T**是决定因素，**T**不包含码

BCNF（续）

❖ 解决方法：将**STJ**分解为二个关系模式：

$ST(S, T) \in BCNF, TJ(T, J) \in BCNF$

没有任何属性对码是部分函数依赖或传递函数依赖

3NF与BCNF的关系

❖ $R \in \text{BCNF} \xrightleftharpoons[\text{不必要}]{\text{充分}} R \in \text{3NF}$

❖ 如果 $R \in \text{3NF}$ ，且 R 只有一个候选码

$R \in \text{BCNF} \xrightleftharpoons[\text{必要}]{\text{充分}} R \in \text{3NF}$

复 习

❖ 函数依赖

1. 平凡的函数依赖，非平凡的函数依赖
2. 部分函数依赖，完全函数依赖
3. 传递函数依赖

❖ 范式

1. 第一范式 (1NF)
2. 第二范式 (2NF)
3. 第三范式 (3NF)
4. 扩展的第三范式 (BCNF)



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

练习

假设某大型连锁超市数据库中有一关系模式R(超市编号, 商品编号, 销售价格, 部门代码, 超市经理)

假设：每个超市可以有多种商品，每种商品可以在多个超市出售；每个超市的每种商品只在一个部门销售；每个超市只有一个经理；每个超市的每种商品只有一个销售价格。

- 1、请指出此关系模式的候选码，给出基本函数依赖，判断此关系模式是第几范式？
- 2、如果不是第三范式，请将其规范化为第三范式的关系模式，并指出分解后的各个关系模式的主码和外码。

作业



P101 6、7、8

Theory of Database