



第二篇 设计与应用开发篇

第六章 关系数据理论

福州大学数计学院

程 烨

chengye@fzu.edu.cn

2013年4月25日

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

- 针对具体问题，如何构造一个适合于它的**数据模式(设计数据库的全局逻辑结构)**
- 数据库逻辑设计的工具——
关系数据库的规范化理论

6.1 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

一、概念回顾

- ❖ 关系
- ❖ 关系模式
- ❖ 关系数据库
- ❖ 关系数据库的模式

一、概念回顾

关系：描述实体型及实体型间的联系。

从形式上看，它是一张二维表，是所涉及属性值域上的笛卡尔积的一个子集。

关系模式：是对关系静态特征(逻辑结构)的描述。

一、概念回顾

关系数据库：基于关系模型的数据库，利用关系来描述现实世界。

- 从形式上看，它由一组**关系**组成。

关系数据库的模式：定义这组关系的**关系模式**的集合。

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: 属性间数据的依赖关系集合

三、什么是数据依赖

1. 完整性约束的表现形式

- ❖ 限定属性取值范围：例如学生成绩必须在0-100之间
- ❖ 定义属性值间的相互关连（主要体现于值的相等与否），这就是数据依赖，它是数据库模式设计的关键

什么是数据依赖（续）

2. 数据依赖

- ❖ 一个关系内部属性与属性之间的约束关系
- ❖ 现实世界属性间相互联系的抽象
- ❖ 数据内在的性质
- ❖ 语义的体现

什么是数据依赖（续）

3. 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)
- ❖ 连接依赖

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 r 满足 F 时， r 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**)

系主任姓名 (**Mname**)、课程号 (**Cno**)

成绩 (**Grade**)

单一的关系模式：**Student** <U、F>

$U = \{ Sno, Sdept, Mname, Cno, Grade \}$

五、数据依赖对关系模式的影响

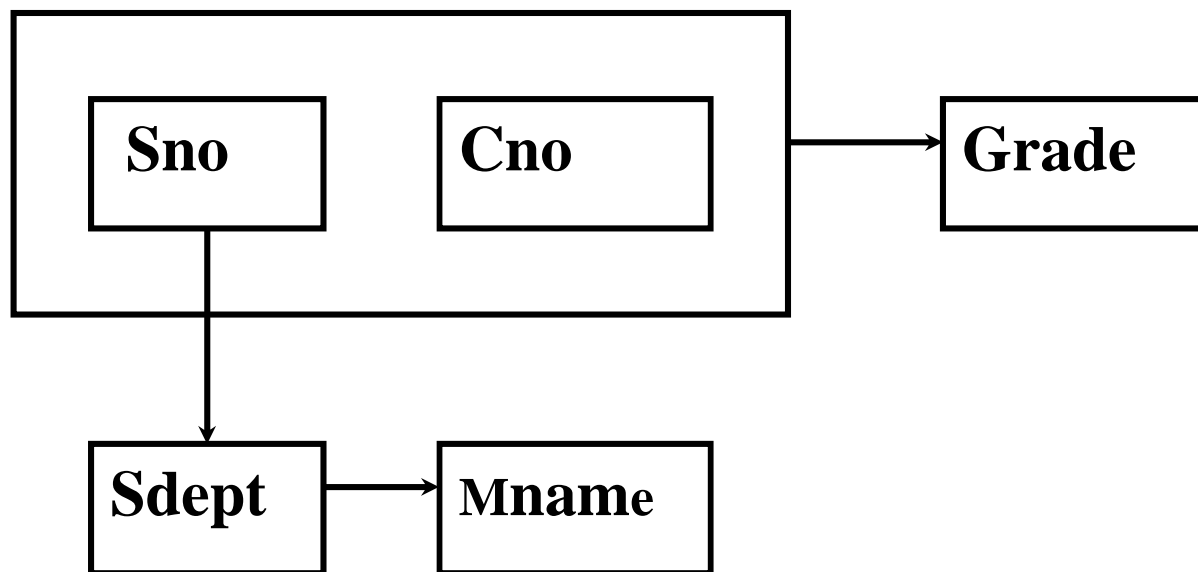
数据库的语义：

1. 一个系有若干学生，
一个学生只属于一个系；
2. 一个系只有一名主任；
3. 一个学生可以选修多门课程，
每门课程有若干学生选修；
4. 每个学生所学的每门课程都有一个成绩。

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:

$$F = \{ \text{Sno} \rightarrow \text{Sdept}, \text{Sdept} \rightarrow \text{Mname}, \\ (\text{Sno}, \text{Cno}) \rightarrow \text{Grade} \}$$



关系模式Student<U, F>中存在的问题

1. 数据冗余太大 (参见表6.1)

例：系主任的姓名在每个元组中重复出现

2. 更新异常(参见表6.1)

- 数据冗余, 更新数据时, 维护数据完整性代价大。

例：某系更换系主任后, 系统必须修改与该系学生有关的每一个元组。

关系模式Student<U, F>中存在的问题

3. 插入异常_该插的数据插不进去(参见表6.1)

例，如果一个系刚成立，尚无学生，无法把这个系及其系主任的信息存入数据库。

4. 删除异常_不该删除的数据不得不删(参见表6.1)

例，如果某个系的学生全部毕业了，我们在删除该系学生信息的同时，把这个系及其系主任的信息也丢掉了。

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：

$S(Sno, Sdept, Sno \rightarrow Sdept)$;

$SC(Sno, Cno, Grade, (Sno, Cno) \rightarrow Grade)$;

$DEPT(Sdept, Mname, Sdept \rightarrow Mname)$

数据依赖对关系模式的影响（续）

关系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。

数据依赖对关系模式的影响（续）

下面我们将讨论：

- 1、如何依据属性之间的数据依赖关系判定关系模式的**规范化程度**（1NF，2NF，3NF，BCNF，4NF）
- 2、属性之间的函数依赖关系的**公理系统**
- 3、如何将某个关系模式**转换**为满足某种规范化的形式。

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

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

- ❖ 函数依赖与互为函数依赖
- ❖ 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、函数依赖与互为函数依赖

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 $X \rightarrow Y$ 。

一、函数依赖与互为函数依赖

函数依赖只能根据语义来确定，且函数依赖指关系模式R上所有关系实例都要满足的条件。

一些记号和术语：

1、若 $x \rightarrow y$ ，称 x 为这函数依赖的决定因素。

2、若 $x \rightarrow y$ ， $y \rightarrow x$ ，则记作 $x \leftrightarrow y$ ， x ， y 互为函数地依赖。

3、 y 不函数依赖于 x ，记作 $x \nrightarrow y$

一、函数依赖与互为函数依赖

例：S (S#, SN, SD, SA) 中，
S# \rightarrow SD, S# \rightarrow SA, S# 是决定因素
S# \leftrightarrow SN (不重名)

例：SC (S#, C#, G) 中
(S#, C#) \rightarrow G,
S# \nrightarrow G, C# \nrightarrow G

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 **$Y \not\subseteq X$** ，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： **$(Sno, Cno) \rightarrow Grade$**

平凡函数依赖： **$(Sno, Cno) \rightarrow Sno$**

$(Sno, Cno) \rightarrow Cno$

（通常，我们总是讨论非平凡的函数依赖）

三、完全函数依赖与部分函数依赖

定义6.2 在 $R(U)$ 中, 如果 $X \rightarrow Y$, 并且对于 X 的任何一个真子集 X' , 都有 $X' \not\rightarrow Y$, 则称 Y 对 X 完全函数依赖, 记作 $X \xrightarrow{F} Y$ 。

若 $X \rightarrow Y$, 但 Y 不完全函数依赖于 X , 则称 Y 对 X 部分函数依赖, 记作 $X \xrightarrow{P} Y$ 。

完全函数依赖与部分函数依赖（续）

[例1] 中 $(Sno, Cno) \xrightarrow{F} Grade$ 是完全函数依赖，

$(Sno, Cno) \xrightarrow{P} Sdept$ 是部分函数依赖

因为 $Sno \rightarrow Sdept$ 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 $X \rightarrow Y$, $(Y \not\subseteq X)$, $Y \not\rightarrow X$ $Y \rightarrow Z$, 则称 Z 对 X 传递函数依赖。

记为: $X \xrightarrow{\text{传递}} Z$

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

例: 在关系 $Std(Sno, Sdept, Mname)$ 中, 有:

$Sno \rightarrow Sdept, Sdept \rightarrow Mname$

$Mname$ 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

定义6.4 设 K 为 $R\langle U, F \rangle$ 中的属性或属性组合。若 $K \xrightarrow{F} U$ ，
则 K 称为 R 的**候选码**（**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(Sno, Sdept, Sage)**, 单个属性**Sno**是码,
SC (Sno, Cno, 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 在 $R(U_1)$ 中, $S(U_2)$ 中, 若 $K \subseteq U_1$, $K \subseteq U_2$, K 是 S 的主码, 且 K 不是 R 的主码, 则称 K 为 R 的**外码**。

如 Sno 是关系模式 $S(\underline{Sno}, Sdept, Sage)$ 的主码, 则 Sno 是关系模式 $SC(\underline{Sno}, \underline{Cno}, Grade)$ 的外码

❖ 主码与外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

关系数据库中的关系，按其属性间满足不同程度的要求，定义了不同级别的范式。

某级范式是满足某级要求的关系模式的集合。

满足最基本要求的称为第一范式关系模式（1NF），在1NF基础上，进一步再满足一些要求，形成第二范式关系模式（2NF），以此类推，目前已提出到5NF，其中3NF又有一种增强范式称为BCNF。

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**最高**为第n范式，可简记为 $R \in nNF$ 。

6.2.3 范式

把一个给定的关系模式转化为某种范式的过程称为关系模式的规范化过程，简称规范化。

以函数依赖为基础的关系模式的范式主要有四种：
1NF，2NF，3NF，BCNF范式。

满足这些范式条件的关系模式可以从不同程度上避免前面提到的插入、删除、更新、冗余问题。

第四范式涉及多值依赖引起的问题。

第五范式解决涉及连接依赖引起的问题。

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

6.2.4 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 **$R \in 1NF$**

- ❖ 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

2NF (续)

例：关系模式 $SLC(S\#, SD, SL, C\#, G)$

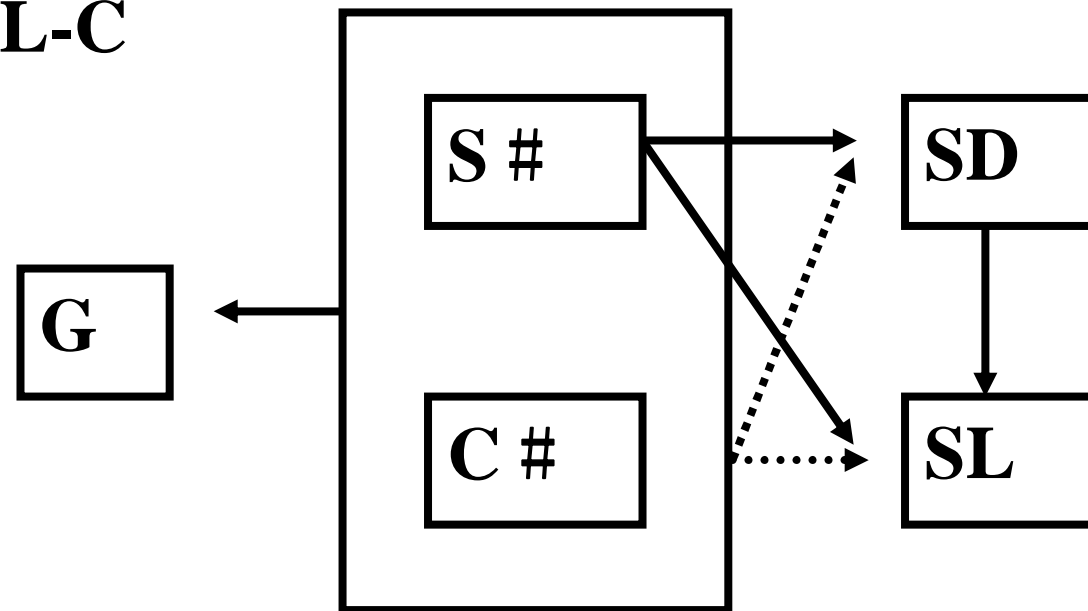
SL为学生住处，假设每个系的学生住在同一个地方。

■ 函数依赖包括：

$$\{(S\#, C\#) \xrightarrow{f} G; S\# \rightarrow SD; (S\#, C\#) \xrightarrow{P} SD; \\ SD \rightarrow SL; (S\#, C\#) \xrightarrow{P} SL\}$$

2NF (续)

S-L-C



- ❖ **SLC**的码为(**S#**, **C#**)
- ❖ **SLC**满足第一范式。
- ❖ 非主属性**SD**和**SL**部分函数依赖于码(**S#**, **C#**)

SLC不是一个好的关系模式（续）

SLC (S#, SD, SL, C#, G) 满足第一范式, 但不是一个好的关系模式

- 原因

SD、SL部分函数依赖于码 (S#, C#) 。

- 解决方法

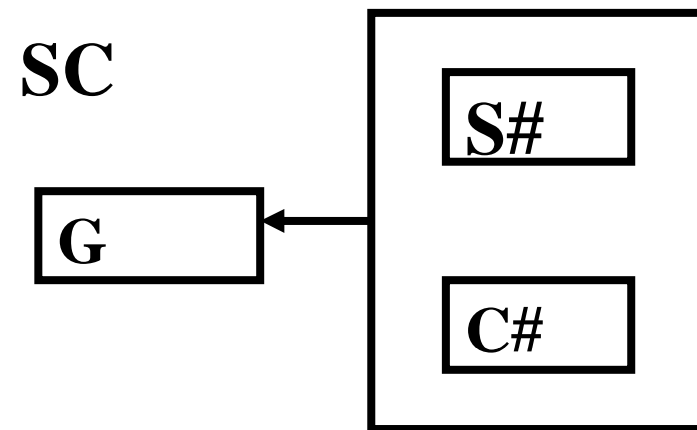
SLC分解为两个关系模式, 以消除这些部分函数依赖。

SLC不是一个好的关系模式（续）

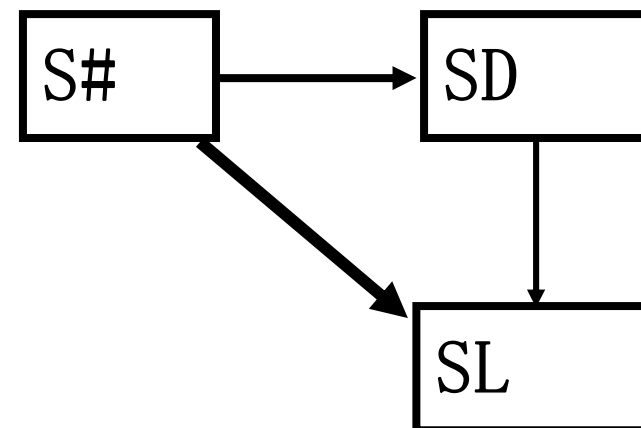
SLC (S#, SD, SL, C#, G)

分解为两个关系模式:

SC (S#, C#, G)



SL




2NF (续)

❖ 2NF的定义

定义6.6 若 $R \in 1NF$ ，且每一个非主属性完全函数依赖于码，则 $R \in 2NF$ 。

关系模式SLC ($S\#, SD, SL, C\#, G$) $\in 1NF$; $\notin 2NF$,

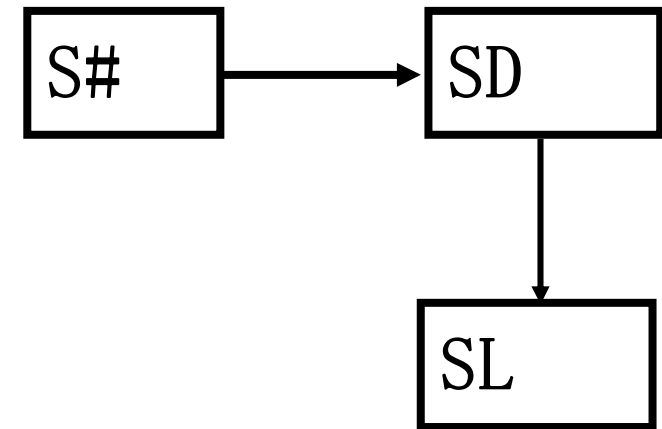
SLC  $\begin{matrix} SC (\underline{S\#}, \underline{C\#}, G) \\ SL (\underline{S\#}, SD, SL) \end{matrix} \in 2NF$

2NF (续)

SL

例：2NF关系模式SL(S#, SD, SL)中
函数依赖：{S#→SD, SD→SL}

SL传递函数依赖于S#，即SL
中存在非主属性对码的传递函数
依赖。



■ 解决方法

将SL分解为两个关系模式，以消除传递函数依赖：

SD (S#, SD), SD的码为S#,

DL (SD, SL), DL的码为SD。

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 $R\langle U, F \rangle$ 中若不存在这样的码

X 、属性组 Y 及非主属性 Z ($Z \not\subseteq Y$), 使得 $X \rightarrow Y$,

$Y \rightarrow Z$ 成立, $Y \not\rightarrow X$,

则称 $R\langle U, F \rangle \in 3NF$ 。

3NF (续)

若 $R \in 2NF$ ，且它的任何一个非主属性都不传递地依赖于任何候选码，则 $R \in 3NF$ 。

即：若 $R \in 3NF$ ，则非主属性不部分或传递依赖于任何候选码。

例： $SL (\underline{S\#}, SD, SL) \in 2NF$;

\because 非主属性 SL 传递地依赖于主键 $S\#$, $\therefore SL \notin 3NF$

解决方法： $SL \longrightarrow \begin{matrix} SD (\underline{S\#}, SD) \\ DL (\underline{SD}, SL) \end{matrix} \in 3NF$

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

6.2.6 BC范式 (BCNF)

❖ 定义6.8 关系模式 $R\langle U, F \rangle \in 1NF$, 若 $X \rightarrow Y$ 且 $Y \not\subseteq X$ 时 X 必含有码, 则 $R\langle U, F \rangle \in BCNF$ 。

❖ 等价于: 每一个决定属性因素都包含码

BCNF（续）

由BCNF定义可得出以下**结论**：

- A. 所有**非主属性**都完全函数依赖于**每个候选码**
- B. 所有**主属性**完全依赖于每个**不包含它的候选码**
- C. **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7：在关系模式SJP (S, J, P) 中，S表示学生，J表示课程，P表示名次（无并列）。

■ F: $\{(S, J) \rightarrow P, (J, P) \rightarrow S\}$

候选键：(S, J), (J, P), S, J, P都是主属性

没有非主属性部分依赖或传递依赖于候选键

$\therefore \text{STJ} \in 3\text{NF}$

\because 对F中任意一个 $X \rightarrow Y$, X都是候选键,

$\therefore \text{STJ} \in \text{BCNF}$

BCNF (续)

例8：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。

- 每一教师只教一门课。每门课可以多位教师授课，某一学生选定某门课，就确定了一位授课教师。某个学生选修某个教师的课就确定了所选课的名称： $(S, J) \rightarrow T$, $(S, T) \rightarrow J$, $T \rightarrow J$

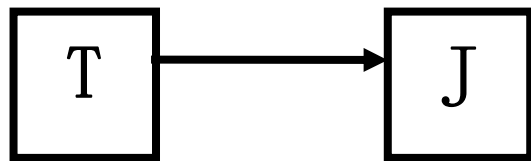
BCNF (续)

- ∵ (S, J) 和 (S, T) 都是候选码
- ∴ S、T、J 都是主属性。
- ∵ 没有非键属性部分依赖或传递依赖于候选键
- ∴ $STJ \in 3NF$
- ∵ $T \rightarrow J$, T 是决定属性集, T 不是候选码
- ∴ $STJ \notin BCNF$

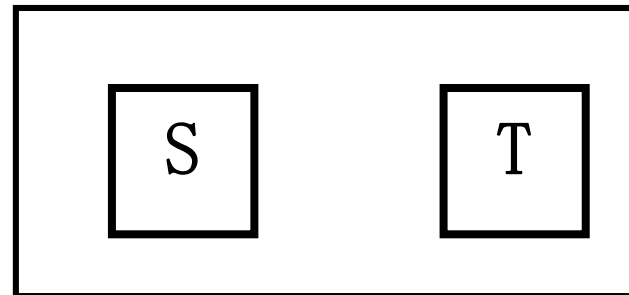
BCNF (续)

解决方法：将STJ分解为二个关系模式：

$TJ(\underline{T}, J) \in \text{BCNF}$, $ST(\underline{S}, \underline{T}) \in \text{BCNF}$



TJ



ST

没有任何属性对码存在部分函数依赖和传递函数依赖。

3NF与BCNF的关系

$$\diamond R \in \text{BCNF} \begin{array}{c} \xrightarrow{\text{充分}} \\ \xleftarrow{\text{不必要}} \end{array} R \in \text{3NF}$$

❖ 如果 $R \in \text{3NF}$ ，且 R 只有一个候选码

$$R \in \text{BCNF} \begin{array}{c} \xrightarrow{\text{充分}} \\ \xleftarrow{\text{必要}} \end{array} R \in \text{3NF}$$

❖ 如果 R 的候选码是全码, 则 $R \in \text{BCNF}$

若一个关系模式 $\in \text{BCNF}$ ，在函数依赖范畴内，它已达到最高规范化程度，已消除了插入、删除的异常。

例子

写出下列各关系模式的函数依赖集F，指出主键，并说明该关系模式最高属第几范式？

BR（借书证号R#，读者姓名RN，单位DW，电话TEL，书号B#，书名BN，出版社BP，出版社地址BPA，借阅期DATE）

$$F = \{R\# \rightarrow RN, R\# \rightarrow DW, R\# \rightarrow TEL, \\ B\# \rightarrow BN, B\# \rightarrow BP, BP \rightarrow BPA, \\ (R\#, B\#) \rightarrow DATE\}$$

例子

$F = \{R\# \rightarrow RN, R\# \rightarrow DW, R\# \rightarrow TEL,$
 $B\# \rightarrow BN, B\# \rightarrow BP, BP \rightarrow BPA,$
 $(R\#, B\#) \rightarrow DATE\}$

主键: $(R\#, B\#)$

非键属性 $\{RN, DW, TEL, BN, BP, BPA, DATE\}$

\because 存在非键属性对键的部分函数依赖关系

$\therefore BR \in 1NF$

1NF-BCNF小结:

1NF



消除非主属性对候选码的部分函数依赖

2NF



消除非主属性对候选码的传递函数依赖

3NF



消除主属性对候选码的部分或传递函数依赖

BCNF

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。

多值依赖（续）

❖ 非规范化关系

课 程 C	教 员 T	参 考 书 B
物理	$\left\{ \begin{array}{l} \text{李 勇} \\ \text{王 军} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{普通物理学} \\ \text{光学原理} \\ \text{物理习题集} \end{array} \right\}$
数学	$\left\{ \begin{array}{l} \text{李 勇} \\ \text{张 平} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{数学分析} \\ \text{微分方程} \\ \text{高等代数} \end{array} \right\}$
计算数学 ⋮	$\left\{ \begin{array}{l} \text{张 平} \\ \text{周 峰} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{数学分析} \\ \vdots \end{array} \right\}$

多值依赖（续）

用二维表表示Teaching

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

...

...

...

多值依赖（续）

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

❖ $\therefore \text{Teaching} \in \text{BCNF}$

存在的问题：数据冗余问题。

插入一个教师要插入多个元组，删一本参考书，要删多个元组，原因是**多值依赖**的数据依赖引起的。

根本原因：教员和参考书之间的联系不是直接的联系，而是间接的联系。把有间接联系的属性放在一个模式中就会产生冗余和异常现象。

多值依赖（续）

❖ 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 $X \twoheadrightarrow Y$ 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), $C \twoheadrightarrow T$, $C \twoheadrightarrow B$

多值依赖（续）

[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仓库的所有商品
- 每种商品被所在仓库的所有保管员保管

多值依赖（续）

W	S	C
W1	S1	C1
W1	S1	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

关系模式 **WSC** (**W**, **S**, **C**) , $W \twoheadrightarrow S$, $W \twoheadrightarrow C$

多值依赖（续）

函数依赖可以看成是多值依赖的特殊情况。

若 $X \twoheadrightarrow Y$ ，而 Z 为空集（即 $U=X+Y$ ），则称 $X \twoheadrightarrow Y$ 为平凡的多值依赖。

若 $X \twoheadrightarrow Y$ ，而 Z 为非空集，则称 $X \twoheadrightarrow Y$ 为非平凡的多值依赖。

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

6.2.8 4NF

- ❖ 定义6.10 关系模式 $R \in 1NF$ ，如果对于 R 的每个非平凡多值依赖 $X \twoheadrightarrow Y$ ($Y \subsetneq X$)， X 都含有码，则 $R \in 4NF$ 。
- ❖ 如果 $R \in 4NF$ ， 则 $R \in BCNF$
 - 不允许有非平凡且非函数依赖的多值依赖
 - 允许的非平凡多值依赖是函数依赖

4NF (续)

例: **Teaching(C,T,B) \notin 4NF**

存在非平凡的多值依赖 **$C \twoheadrightarrow T$, $C \twoheadrightarrow B$**

且**C**不是码

- 用投影分解法把**Teaching**分解为如下两个关系模式:

$CT(C, T) \in 4NF$

$CB(C, B) \in 4NF$

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

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

6.3 数据依赖的公理系统

- 1 Armstrong公理系统(重点)
- 2 函数依赖集的闭包
- 3 属性集闭包的计算(重点)
- 4 函数依赖集的等价性
- 5 最小函数依赖集(重点)

6.3 数据依赖的公理系统

❖ 逻辑蕴含

定义6.11 对于满足一组函数依赖 F 的关系模式 $R \langle U, F \rangle$ ，其任何一个关系 r ，若函数依赖 $X \rightarrow Y$ 都成立，（即 r 中任意两元组 t, s ，若 $t[X] = s[X]$ ，则 $t[Y] = s[Y]$ ），则称 F 逻辑蕴含 $X \rightarrow Y$

为了求得给定关系模式的候选键，为给一组函数依赖求得所蕴涵的所有函数依赖，就需要一套推理规则。

1. Armstrong公理系统

对 $R(U, F)$ ，有以下的推理规则：

A1. 自反律：若 $Y \subseteq X \subseteq U$ ，则 F 蕴涵 $X \rightarrow Y$

注意：由自反律所得到的函数依赖均是平凡的函数依赖，自反律的使用并不依赖于 F 。

A2. 增广律：若 F 蕴涵 $X \rightarrow Y$ ， $Z \subseteq U$ ，则 F 蕴涵 $XZ \rightarrow YZ$

A3. 传递律：若 F 蕴涵 $X \rightarrow Y$ 和 $Y \rightarrow Z$ ，则 F 蕴涵 $X \rightarrow Z$

可以证明Armstrong推理规则的正确性 P183。

导出规则

1. 根据 **A1**, **A2**, **A3** 这三条推理规则可以得到下面三条推理规则:

- **合并规则**: 由 $X \rightarrow Y$, $X \rightarrow Z$, 有 $X \rightarrow YZ$.
- **伪传递规则**: 由 $X \rightarrow Y$, $WY \rightarrow Z$, 有 $XW \rightarrow Z$.
- **分解规则**: 由 $X \rightarrow Y$ 及 $Z \subseteq Y$, 有 $X \rightarrow Z$.

导出规则

2.由合并规则： 如果 $X \rightarrow Y$, $X \rightarrow Z$, 则 $X \rightarrow YZ$

和分解规则： 如果 $X \rightarrow YZ$, 则 $X \rightarrow Y$ 、 $X \rightarrow Z$

可得到引理**6.1**

引理6.1 $X \rightarrow A_1 A_2 \dots A_k$ 成立的充分必要条件是
 $X \rightarrow A_i$ 成立 ($i=1, 2, \dots, k$)

例子

设 $F = \{AB \rightarrow E, AC \rightarrow F, AD \rightarrow BF, B \rightarrow C, C \rightarrow D\}$

证 $AC \rightarrow F$ 是冗余的。

证： $AC \rightarrow F$ 可由其它函数依赖推出

$\because C \rightarrow D \quad \therefore AC \rightarrow AD$ 增广律

$\because AC \rightarrow AD, AD \rightarrow BF \quad \therefore AC \rightarrow BF$ 传递律

$\therefore AC \rightarrow F$ 分解规则或引理

证毕

2 函数依赖集的闭包

定义**6.12**：对于一个 $R(U, F)$ ，由 F 逻辑蕴涵的**所有**函数依赖称为 **F 的闭包**，记作 F^+

可以证明Armstrong公理系统是**有效而且完备**的。

有效性是指由 F 出发依据公理推出的每个函数依赖一定在 F^+ 中；

完备性指的是 F^+ 中的每个函数依赖一定可以由 F 出发根据公理推导出来。

F的闭包

$$F = \{X \rightarrow Y, Y \rightarrow Z\}$$

$$F^+ = \left\{ \begin{array}{llll} X \rightarrow \varnothing, & Y \rightarrow \varnothing, & Z \rightarrow \varnothing, & XY \rightarrow \varnothing, \quad XZ \rightarrow \varnothing, \quad YZ \rightarrow \varnothing, \quad XYZ \rightarrow \varnothing, \\ X \rightarrow X, & Y \rightarrow Y, & Z \rightarrow Z, & XY \rightarrow X, \quad XZ \rightarrow X, \quad YZ \rightarrow Y, \quad XYZ \rightarrow X, \\ X \rightarrow Y, & Y \rightarrow Z, & & XY \rightarrow Y, \quad XZ \rightarrow Y, \quad YZ \rightarrow Z, \quad XYZ \rightarrow Y, \\ X \rightarrow Z, & Y \rightarrow YZ, & & XY \rightarrow Z, \quad XZ \rightarrow Z, \quad YZ \rightarrow YZ, \quad XYZ \rightarrow Z, \\ X \rightarrow XY, & & & XY \rightarrow XY, \quad XZ \rightarrow XY, \quad XYZ \rightarrow XY, \\ X \rightarrow XZ, & & & XY \rightarrow YZ, \quad XZ \rightarrow XZ, \quad XYZ \rightarrow YZ, \\ X \rightarrow YZ, & & & XY \rightarrow XZ, \quad XZ \rightarrow XY, \quad XYZ \rightarrow XZ, \\ X \rightarrow XYZ, & & & XY \rightarrow XYZ, \quad XZ \rightarrow XYZ, \quad XYZ \rightarrow XYZ \end{array} \right\}$$

F的闭包 F^+ 计算是一个NP完全问题

2 函数依赖集的闭包（续）

实际上，函数依赖是关键码概念的推广。

设关系模式R的属性集是U，X是U的一个子集。如果 $X \rightarrow U$ 在R上成立，那么称X是R的一个**超键**。如果 $X \rightarrow U$ 在R上成立，但对于X的任一真子集 X_1 都有 $X_1 \rightarrow U$ 不成立，那么称X是R上的一个**候选键**。

2 函数依赖集的闭包（续）

在实际使用中，经常要判断能否从已知的FD集 F 推导出FD $X \rightarrow Y$ ，那么可先求出 F 的闭包 F^+ ，然后再看 $X \rightarrow Y$ 是否在 F^+ 中。

但是从 F 求 F^+ 是一个复杂且困难的问题（NP完全问题，指数级问题）。如果求属性集闭包的概念，将使判断问题化为多项式级时间问题。

3 属性集闭包的计算

定义6.13 设 F 为属性集 U 上的一组函数依赖,

$X \subseteq U$, $X_F^+ = \{ A/X \rightarrow A \text{ 能由 } F \text{ 根据 Armstrong 公理导出} \}$, X_F^+ 称为属性集 X 关于函数依赖集 F 的闭包

说明: X_F^+ 包含了所有由 F 根据公理系统, 属性 X 可确定的属性集。

3 属性集闭包的计算

设有关系模式 $R(U, F)$ ， $U=\{A, B, C\}$ ， $F=\{A\rightarrow B, B\rightarrow C\}$

如果 $X=\{A\}$ ，则：

$$X_F^+ = \{A, B, C\}$$

如果 $X=\{B\}$ ，则

$$X_F^+ = \{B, C\}$$

如果 $X=\{C\}$ ，则

$$X_F^+ = \{C\}$$

关于闭包的引理

引理6.2: 设 F 为属性集 U 上的一组函数依赖, $X, Y \subseteq U$, $X \rightarrow Y$ 能由 F 根据公理推出的**充要条件**是 $Y \subseteq X_F^+$

$$X \rightarrow Y \in F^+ \Leftrightarrow Y \subseteq X_F^+$$

这样, 判断 $X \rightarrow Y$ 是否可由 F 根据公理推出的问题, 就转化为求出 X_F^+ , 判断 Y 是否为 X_F^+ 的子集的问题。

求闭包的算法

算法6.1 求属性集 X ($X \subseteq U$) 关于 U 上的函数依赖集 F 的闭包 X_F^+

输入: X, F 输出: X_F^+

步骤:

- (1) 令 $X^{(0)} = X, i=0$
- (2) 求 B , 这里 $B = \{ A \mid (\exists V)(\exists W)(V \rightarrow W \in F \wedge V \subseteq X^{(i)} \wedge A \in W) \}$;
- (3) $X^{(i+1)} = B \cup X^{(i)}$
- (4) 判断 $X^{(i+1)} = X^{(i)}$ 吗?
- (5) 若相等或 $X^{(i)} = U$, 则 $X^{(i)}$ 就是 X_F^+ , 算法终止。
- (6) 若否, 则 $i=i+1$, 返回第(2)步。

属性集的闭包

例：已知R的属性集 $U=\{A, B, C, D, E\}$

$F=\{AB\rightarrow C, AC\rightarrow B, B\rightarrow D, C\rightarrow E, EC\rightarrow B\}$,

求 $(AB)_F^+$

解：1)、 $X^{(0)}=\{A, B\}$, $i=0$.

2)、求B, 找左边为A, B或AB的函数依赖。

得到： $\{AB\rightarrow C, B\rightarrow D\} \therefore B=\{C, D\}$

$X^{(1)}=B \cup X^{(0)}=\{A, B, C, D\}$

$\therefore X^{(1)} \neq X^{(0)} \therefore i=1$ 。

属性集的闭包

3)、求B, 找左边为ABCD子集的函数依赖。

得到: $\{AC \rightarrow B, C \rightarrow E\} \therefore B = \{BE\}$

$X^{(2)} = B \cup X^{(1)} = \{A, B, C, D, E\} = U$

$\therefore (AB)_F^+ = \{A, B, C, D, E\}$

3 属性集闭包的计算

设有关系模式R (A, B, C, D) , 其上的函数依赖集:

$F = \{A \rightarrow C, C \rightarrow A, B \rightarrow AC, D \rightarrow AC\}$, 试计算: $(AD)^+$

令 $X = \{AD\}$,

$X(0) = \{AD\}$,

$X(1) = \{ACD\}$,

$X(2) = \{ACD\} = X(1)$,

故 $(AD)^+ = \{ACD\}$

4. 函数依赖集等价

定义： 设 F 和 G 是两个函数依赖集：

- ①如果 $F \subseteq G^+$ ，则称 G 是 F 的一个覆盖，或称 G 覆盖 F ；
- ②如果 $F \subseteq G^+$ 和 $G \subseteq F^+$ 同时成立，即 $F^+ = G^+$ ，则称 F 和 G 等价。

4. 函数依赖集等价

引理： 设G和F是两个函数依赖集，F与G等价的**充要条件**是 $F \subseteq G^+$ 且 $G \subseteq F^+$

要判断 $F \subseteq G^+$ ，只须对F中的**每个**函数依赖 $X \rightarrow Y$ ，求出 X_G^+ ，判定Y是否是 X_G^+ 的子集。

同样可判定 $G \subseteq F^+$

练习题

假设有一个名为**参加**的关系，该关系有属性：职工（职工名）、工程（工程名）、时数（花费在工程上的小时数）和工资（职工在该工程上的工资）；一个**参加关系的元组**描述一个职工花费在一个工程上的时数和工资；另外，一个职工可以参加多个工程，多个职工可以参加同一个工程。请找出这个关系中的所有函数依赖；确定这个关系的关键字；

练习题

设用A、B、C、D分别代表属性职工、工程、时数和工资。 $R(A, B, C, D)$

1、这个关系中的所有函数依赖：

$$F = \{AB \rightarrow C, AB \rightarrow D\}$$

2、关系的关键字：

AB

3、R最高属 ?NF:

BCNF

练习题

依据下列各关系模式的函数依赖集F，指出主键，并说明该关系模式最高属第几范式？

1、 $R(A, B, C, D)$, $F\{B \rightarrow D, AB \rightarrow C\}$

主键AB，非主属性C, D

\because 存在非主属性对键的部分依赖

$\therefore R \in 1NF$

练习题

依据下列各关系模式的函数依赖集F，指出主键，并说明该关系模式最高属第几范式？

2、 $R(A, B, C, D, E)$, $F\{AB \rightarrow CE, E \rightarrow AB, C \rightarrow D\}$

候选码AB, E, 非主属性C, D

$\because AB \rightarrow C, C \rightarrow D$

\therefore 非主属性D对码AB存在传递依赖

$\therefore R \in 2NF$

练习题

依据下列各关系模式的函数依赖集F，指出主键，并说明该关系模式最高属第几范式？

3、 $R(A, B, C, D)$, $F\{B \rightarrow D, D \rightarrow B, AB \rightarrow C\}$

候选码AB, AD , 非主属性C

$\because B \rightarrow D$, B不是候选码

$\therefore R \in 3NF$

练习题

依据下列各关系模式的函数依赖集F，指出主键，并说明该关系模式最高属第几范式？

4、R (A, B, C) , F {A→B, B→A , A→C}

候选码：A, B 非主属性C

∵ 对任意一个 $X \rightarrow Y$, X都是候选码

∴ $R \in \text{BCNF}$

6.3 数据依赖的公理系统

- 1 Armstrong公理系统
- 2 函数依赖集的闭包
- 3 属性集闭包的计算
- 4 函数依赖集的等价性
- 5 最小函数依赖集

5. 最小依赖集

定义6.15 如果函数依赖集**F**满足下列条件，则称**F**为一个极小函数依赖集。亦称为最小依赖集或最小覆盖。

(1) **F**中任一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。

5. 最小依赖集

定义的分析：

条件（1）保证在函数依赖的右部没有多余的属性

（2）保证F中没有多余的函数依赖

（3）保证在函数依赖的左部没有多余的属性

定理： 每个函数依赖集合F均等价于一个最小函数依赖集合 F_m

最小依赖集

[例] 关系模式 $S\langle U, F \rangle$, 其中:

$U = \{ \text{Sno}, \text{Sdept}, \text{Mname}, \text{Cno}, \text{Grade} \},$

$F = \{ \text{Sno} \rightarrow \text{Sdept}, \text{Sdept} \rightarrow \text{Mname},$

$(\text{Sno}, \text{Cno}) \rightarrow \text{Grade} \}$

设 $F' = \{ \text{Sno} \rightarrow \text{Sdept}, \text{Sno} \rightarrow \text{Mname}, \text{Sdept} \rightarrow \text{Mname},$

$(\text{Sno}, \text{Cno}) \rightarrow \text{Grade}, (\text{Sno}, \text{Sdept}) \rightarrow \text{Sdept} \}$

F 是最小覆盖, 而 F' 不是。

因为: $F' - \{ \text{Sno} \rightarrow \text{Mname} \}$ 与 F' 等价

$F' - \{ (\text{Sno}, \text{Sdept}) \rightarrow \text{Sdept} \}$ 也与 F' 等价

极小化过程

定理6.3 每一个函数依赖集 F 均等价于一个极小函数依赖集 F_m 。此 F_m 称为 F 的最小依赖集。

证明: 构造性证明, 找出 F 的一个最小依赖集。

极小化过程

- (1) 逐一检查 F 中各函数依赖 $FD_i: X \rightarrow Y$, 若 $Y=A_1A_2 \cdots A_k$, $k > 2$, 则用 $\{ X \rightarrow A_j \mid j=1, 2, \cdots, k \}$ 来取代 $X \rightarrow Y$ 。
- (2) 逐一检查 F 中各函数依赖 $FD_i: X \rightarrow A$, 令 $G=F - \{X \rightarrow A\}$, 若 $A \in X_G^+$, 则从 F 中去掉此函数依赖。
- (3) 逐一取出 F 中各函数依赖 $FD_i: X \rightarrow A$, 设 $X=B_1B_2 \cdots B_m$, 逐一考查 B_i ($i=1, 2, \cdots, m$), 若 $A \in (X-B_i)_F^+$, 则以 $X-B_i$ 取代 X 。

极小化过程（续）

例1：求 $F = \{A \rightarrow BC, B \rightarrow AC, C \rightarrow A\}$ 的最小函数依赖集合 F_m

解1：1) 先分解右端，

$$F = \{A \rightarrow B, A \rightarrow C, B \rightarrow A, B \rightarrow C, C \rightarrow A\}$$

2) 从左向右逐一检查F中的函数依赖是否多余：

设 $G = F - \{A \rightarrow B\}$ ， $A_G^+ = AC$ ，不包含B， $A \rightarrow B$ 保留

设 $G = F - \{A \rightarrow C\}$ ， $A_G^+ = ABC$ ，包含C， $A \rightarrow C$ 删除，

$$\text{令 } F = G = \{A \rightarrow B, B \rightarrow A, B \rightarrow C, C \rightarrow A\}$$

设 $G = F - \{B \rightarrow A\}$ ， $B_G^+ = BAC$ ，包含A， $B \rightarrow A$ 删除，

$$\text{令 } F = G = \{A \rightarrow B, B \rightarrow C, C \rightarrow A\}$$

设 $G = F - \{B \rightarrow C\}$ ， $B_G^+ = B$ ，不包含C， $B \rightarrow C$ 保留

设 $G = F - \{C \rightarrow A\}$ ， $C_G^+ = C$ ，不包含A， $C \rightarrow A$ 保留

极小化过程（续）

3) F已满足条件 (3) 左端都为单属性。

$\therefore F_m$ 为 $\{A \rightarrow B, B \rightarrow C, C \rightarrow A\}$

极小化过程（续）

例1：求 $F = \{A \rightarrow BC, B \rightarrow AC, C \rightarrow A\}$ 的最小函数依赖集合 F_m

解2：1) 得 $F = \{A \rightarrow B, A \rightarrow C, B \rightarrow A, B \rightarrow C, C \rightarrow A\}$

2) 从右向左逐一检查F中的函数依赖是否多余：

∵ 消去 $B \rightarrow C$ 后，

再无多余的函数依赖。

∴ F_{m2} 为 $\{A \rightarrow B, A \rightarrow C, B \rightarrow A, C \rightarrow A\}$

这样F就有二个不同的最小函数依赖集。

∴ 最小函数依赖集不是唯一的。

极小化过程（续）

例2：求 $F = \{BE \rightarrow G, BD \rightarrow G, CDE \rightarrow AB, CD \rightarrow A, CE \rightarrow G, BC \rightarrow A, B \rightarrow D, C \rightarrow D\}$ 的最小函数依赖集合 F_m

解：1) 先分解右端，

$$F = \{BE \rightarrow G, BD \rightarrow G, CDE \rightarrow A, CDE \rightarrow B, CD \rightarrow A, CE \rightarrow G, BC \rightarrow A, B \rightarrow D, C \rightarrow D\}$$

2) 逐一检查F中的函数依赖是否多余（过程略）

$$F = \{BD \rightarrow G, CDE \rightarrow B, CD \rightarrow A, B \rightarrow D, C \rightarrow D\}$$

极小化过程（续）

$$F = \{BD \rightarrow G, CDE \rightarrow B, CD \rightarrow A, B \rightarrow D, C \rightarrow D\}$$

3) 逐一检查并消去F中函数依赖左部的多余属性。

$$B_F^+ = BDG, \text{ 包含 } G, B \rightarrow G \text{ 代替 } BD \rightarrow G$$

$$F = \{B \rightarrow G, CDE \rightarrow B, CD \rightarrow A, B \rightarrow D, C \rightarrow D\}$$

$$CE_F^+ = CEBDGA, \text{ 包含 } B, CE \rightarrow B \text{ 代替 } CDE \rightarrow B$$

$$F = \{B \rightarrow G, CE \rightarrow B, CD \rightarrow A, B \rightarrow D, C \rightarrow D\}$$

$$C_F^+ = CDA, \text{ 包含 } A, C \rightarrow A \text{ 代替 } CD \rightarrow A$$

$$F = \{B \rightarrow G, CE \rightarrow B, C \rightarrow A, B \rightarrow D, C \rightarrow D\}$$

$$\therefore F_m = \{B \rightarrow G, CE \rightarrow B, C \rightarrow A, B \rightarrow D, C \rightarrow D\}$$

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

6.4 模式的分解

- 将较低等级的关系模式分解为若干个较高等级的关系模式过程称为**模式分解**。模式分解的方法并**不是唯一的**。
- 只有能够保证分解后的关系模式与原关系模式等价，分解方法才有意义。

关系模式分解的标准

三种模式分解等价的定义：

1. 分解具有无损连接性
2. 分解要保持函数依赖
3. 分解既要保持函数依赖，又要具有无损连接性

模式的分解（续）

定义6.16 关系模式 $R\langle U, F \rangle$ 的一个分解:

$$\rho = \{ R_1\langle U_1, F_1 \rangle, R_2\langle U_2, F_2 \rangle, \dots, R_n\langle U_n, F_n \rangle \}$$

$U = \bigcup_{i=1}^n U_i$, 且不存在 $U_i \subseteq U_j$, F_i 为 F 在 U_i 上的投影

定义6.17 函数依赖集合 $\{X \rightarrow Y \mid X \rightarrow Y \in F^+ \wedge XY \subseteq U_i\}$ 的一个覆盖 F_i 叫作 F 在属性 U_i 上的投影

模式的分解（续）

例：关系R (U, F) , $U = \{S\#, SD, MN\}$, $F = \{S\# \rightarrow SD, SD \rightarrow MN\}$

语义：学生S#在SD系学习，其系主任是MN。

由于存在MN对S#的传递函数依赖，会出现更新异常，插入、删除异常。

1、对 R进行分解： $\rho_1 = \{R_1(S\#), R_2(SD), R_3(MN)\}$

存在问题：A、原先的函数依赖关系不存在。

B、 $R_1 \bowtie R_2 \bowtie R_3$ 为笛卡尔积，元组增加，不具有无损连接性。

模式的分解（续）

例：关系 $R(U, F)$, $U = \{S\#, SD, MN\}$, $F = \{S\# \rightarrow SD, SD \rightarrow MN\}$

语义：学生 $S\#$ 在 SD 系学习，其系主任是 MN 。

2、对 R 进行另一分解： $\rho_2 = \{R_1(S\#, SD), R_2(S\#, MN)\}$

ρ_2 对 R 的分解是可恢复的，具有无损连接性。

存在问题： $SD \rightarrow MN$ 依赖关系丢失，不具函数依赖保持性。

3、对 R 的分解： $\rho_3 = \{R_1(S\#, SD), R_2(SD, MN)\}$

既具无损连接，又保持函数依赖，同时解决插入、删除、更新异常。

模式的分解（续）

关系模式分解的**目标**：无损连接性和函数依赖保持性。

所谓分解的**无损连接性**：若将 $R(U, F)$ 分解为K个子模式： $\rho = \{R_1(U_1, F_1), R_2(U_2, F_2), \dots, R_k(U_k, F_k)\}$ ，这K个子模式进行**连接运算**后，还可恢复为模式R。

所谓分解的**函数依赖保持性**：若将 $R(U, F)$ 分解为K个子模式： $\rho = \{R_1(U_1, F_1), R_2(U_2, F_2), \dots, R_k(U_k, F_k)\}$ ，则 $F_i (0 \leq i \leq K)$ 并集与F等价。

模式的分解（续）

- ❖ 如果一个分解具有无损连接性，则它能够保证不丢失信息
- ❖ 如果一个分解保持了函数依赖，则它可以减轻或解决各种异常情况
- ❖ 分解具有无损连接性和分解保持函数依赖是两个互相独立的标准。具有无损连接性的分解不一定能够保持函数依赖；同样，保持函数依赖的分解也不一定具有无损连接性。

模式的分解（续）

- ❖ 算法**6.2** 判别一个分解的无损连接性。（教材**P189**）
 - ❖ （1）建立一张**n**列**k**行的表。每一列对应一个属性，每一行对应分解中的一个关系模式。若属性 A_j 属于 U_i ，则在**j**列**i**行交叉处填上 a_j ，否则填上 b_{ij} ；
 - ❖ （2）对每一个 FD_i 做下列操作：找到 X_i 所对应的列中具有相同符号的那些行。考察这些行中 I_i 列的元素，若其中有 a_{li} ，则全部改为 a_{li} ；否则全部改为 b_{mli} ； m 是这些行的行号的最小值。

模式的分解（续）

- ❖ 算法**6.2** 判别一个分解的无损连接性。（教材**P189**）
 - ❖ （3）比较扫描前后，表有无变化。无变化，则终止。
 - ❖ 如果某次更改后，有一行成为 $\mathbf{a_1, a_2, \dots, a_n}$ 。则 ρ 具有无损连接性，否则 ρ 不具有无损连接性。
 - ❖ 例：已知 $\mathbf{R\langle U, F \rangle}$, $\mathbf{U=\{A, B, C, D, E\}}$, $\mathbf{F=\{AB \rightarrow C, C \rightarrow D, D \rightarrow E\}}$, \mathbf{R} 的一个分解为 $\mathbf{R_1(A, B, C), R_2(C, D), R_3(D, E)}$ 。（教材**P190**）

模式的分解（续）

例：设有关系模式 $R(ABC)$ ， $\rho = \{AB, AC\}$ 是 R 的一个分解。试分析分别在 $F1 = \{A \rightarrow B\}$ ， $F2 = \{A \rightarrow C, B \rightarrow C\}$ ， $F3 = \{B \rightarrow A\}$ ， $F4 = \{C \rightarrow B, B \rightarrow A\}$ 情况下， ρ 是否具有无损分解和保持FD的分解特性。

F1: 是无损分解且保持FD集的分解。

F2: 是无损分解，但不保持FD集的分解。 $B \rightarrow C$ 丢失。

F3: 是有损分解，但保持FD集的分解。

F4: 是有损分解且不保持FD集的分解。 $C \rightarrow B$ 丢失

分解算法

模式分解算法：

- 1、分解成3NF并保持函数依赖性的分解算法（算法6.3）
- 2、分解成3NF，既有无损连接性又保持函数依赖性算法（算法6.4）
- 3、具有无损连接性的BCNF分解算法（算法6.5）
- 4、具有无损连接性的4NF分解算法（算法6.6）

要求能应用算法进行模式分解。

分解算法

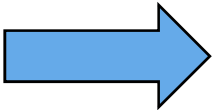
1. 分解成3NF，并保持函数依赖的分解算法：

- 1) 对 $R(U, F)$ 中的 F 进行极小化处理。
- 2) 找出不在 F 中出现的属性，形成一个关系模式。
- 3) 对 F 按具有相同左部的原则分组 F_k ， F_k 所涉及的属性形成属性集 U_k ，若 $U_i \subseteq U_j$ $i \neq j$ ，则去掉 U_i
对每个 U_i 形成子模式 R_i 。

可证明 $R_k \in 3NF$ ，且 ρ 具有函数依赖保持性。

分解算法

例1：图书借阅关系BR（借书证号R#，读者姓名RN，单位DW，电话TEL，书号B#，书名BN，出版社BP，出版社地址BPA，借阅期DATE），将BR分解到3NF并具有函数依赖保持性。

$F = \{$		$\{ R\#, RN, DW, TEL \}$
$R\# \rightarrow RN, R\# \rightarrow DW, R\# \rightarrow TEL,$		$\{ B\#, BN, BP \}$
$B\# \rightarrow BN, B\# \rightarrow BP,$		$\{ BP, BPA \}$
$BP \rightarrow BPA,$		$\{ R\#, B\#, DATE \}$
$(R\#, B\#) \rightarrow DATE \}$		

分解算法

主键 ($R\#$, $B\#$) $BR \in 1NF$

分解: $R_1 (\underline{R\#}, RN, DW, TEL)$

$R_2 (\underline{B\#}, BN, BP)$

$R_3 (\underline{BP}, BPA)$

$R_4 (\underline{R\#}, \underline{B\#}, DATE)$

$R_i \in 3NF$ 且具有函数依赖保持性。

分解算法

例2：关系模式R (U, F)，其中 $F = \{BE \rightarrow G, BD \rightarrow G, CDE \rightarrow AB, CD \rightarrow A, CE \rightarrow G, BC \rightarrow A, B \rightarrow D, C \rightarrow D\}$ ，将R分解到3NF并具有函数依赖保持性。

1、求 $F_m = \{B \rightarrow G, CE \rightarrow B, C \rightarrow A, B \rightarrow D, C \rightarrow D\}$

候选键是CE。

2、分解成三个模式， $R_i \in 3NF$ 且具有函数依赖保持性。

R1: $U_1 = BDG, F_1 = \{B \rightarrow G, B \rightarrow D\}$

R2: $U_2 = ACD, F_2 = \{C \rightarrow A, C \rightarrow D\}$

R3: $U_3 = BCE, F_3 = \{CE \rightarrow B\}$

分解算法

2. 分解成3NF，既有无损连接性又保持函数依赖性算法

(1) 调用算法1产生R的分解 $\rho = \{R_1 \dots R_n\}$

(2) 构造分解 $\tau = \{ R_1, \dots, R_n, R_K \}$ ，其中 R_K 是由R的一个候选键K构成的关系。

分解算法

例：图书借阅关系BR（借书证号R#，读者姓名RN，单位DW，电话TEL，书号B#，书名BN，出版社BP，出版社地址BPA，借阅期DATE），将BR分解到3NF并既具有无损连接性又保持函数依赖性。

$F = \{ R\# \rightarrow RN, R\# \rightarrow DW, R\# \rightarrow TEL, B\# \rightarrow BN, B\# \rightarrow BP, BP \rightarrow BPA, (R\#, B\#) \rightarrow DATE \}$

由算法1得到一组分解：

$R_1 (\underline{R\#}, RN, DW, TEL)$
 $R_2 (\underline{B\#}, BN, BP)$
 $R_3 (\underline{BP}, BPA)$
 $R_4 (\underline{R\#}, \underline{B\#}, DATE)$

$R_i \in 3NF$ 且具有函数依赖保持性。

分解算法

由主键 ($R\#$, $B\#$) , 依据算法2

分解: R_1 ($\underline{R\#}$, RN , DW , TEL) 、

R_2 ($\underline{B\#}$, BN , BP)

R_3 (\underline{BP} , BPA) 、

R_4 ($\underline{R\#}$, $\underline{B\#}$, $DATE$) 、

R_5 ($\underline{R\#}$, $\underline{B\#}$) 因 R_5 的属性包含在 R_4 中, 可删去。

$R_i = \{ R_1, \dots, R_4 \} \in 3NF$

且分解既有无损连接性又保持函数依赖性

分解算法

例2：关系模式R（ABCDE），其中 $F=\{A \rightarrow B, C \rightarrow D\}$ ，将R分解到3NF并既具有无损连接性又保持函数依赖性。

$F_m = \{A \rightarrow B, C \rightarrow D\}$ ，候选键是ACE，

依据算法2分解：R1：U1=AB，F1= $\{A \rightarrow B\}$

R2：U2=CD，F2= $\{C \rightarrow D\}$

R3：U3=ACE，F3= $\{\phi\}$

$R_i = \{ R1, R2, R3 \} \in 3NF$ ，且分解既有无损连接性又保持函数依赖性。

分解算法

3. 具有无损连接性的BCNF分解算法。

输入：关系模式 $R(U, F)$

算法： $\rho = \{R_i\}$;

WHILE ρ 中存在非BCNF的关系模式 DO

 选一个非BCNF模式 $S \in \rho$;

 选 S 的一个违反BCNF要求的函数依赖 $X \rightarrow Y$

 用两个关系模式 $(X \cup Y)$ 和 $(S - Y)$ 取代 ρ 中的 S

ENDWHILE。

分解算法

例1: STJ (学生S, 教师T, 课程J)

每门课程可有多位教师上课、每位教师只上一门课。

F: $\{ (S, J) \rightarrow T, (S, T) \rightarrow J, T \rightarrow J \}$

候选键: (S, J) 、 (S, T) ,

$STJ \in 3NF, \notin BCNF$

对不满足BCNF的依赖关系: $T \rightarrow J$

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

该分解且具有无损连接性。

但损失函数依赖关系: $(S, J) \rightarrow T, (S, T) \rightarrow J$

分解算法

例2: $R(I, J, K, L, M, N)$

$F = \{K \rightarrow M, J \rightarrow N, JK \rightarrow L, N \rightarrow I\}$

分解R使属于BCNF且具有无损连接性。

$\because JK^+ = JKLMNI$ ($J^+ = JNI$, $K^+ = KM$) $\therefore JK$ 是R的主键

1)、取 $N \rightarrow I$, N不是主键

$S_1 = \{N, I\} \in \text{BCNF}$, $F_1 = \{N \rightarrow I\}$

$S_2 = \{JKLMN\}$, $F_2 = \{K \rightarrow M, J \rightarrow N, JK \rightarrow L\}$

S_2 中 $JK^+ = JKLMN$ $\therefore JK$ 是主键

分解算法

2、取 $K \rightarrow M$ $S_{21} = \{K, M\} \in \text{BCNF}$, $F_{21} = \{K \rightarrow M\}$
 $S_{22} = \{J, K, L, N\}$, $F_{22} = \{J \rightarrow N, JK \rightarrow L\}$, 键为JK

3、取 $J \rightarrow N$ $S_{221} = \{J, N\} \in \text{BCNF}$, $F_{221} = \{J \rightarrow N\}$
 $S_{222} = \{J, K, L\} \in \text{BCNF}$, $F_{222} = \{JK \rightarrow L\}$

\therefore 对R的分解 $\rho = \{S_1, S_{21}, S_{221}, S_{222}\} \in \text{BCNF}$
具有无损连接性, 本例函数依赖关系没有损失。

分解算法

例3、R（学生S#，课程号C#，G成绩，TN任课教师，D教师所在系名）， $F\{(S\#, C\#) \rightarrow G, C\# \rightarrow TN, TN \rightarrow D\}$
试将R分解到BCNF且具有无损连接性。

主键（S#，C#）

分解TND（TN，D）， $F1\{TN \rightarrow D\}$

CTN（C#，TN）， $F2\{C\# \rightarrow TN\}$

SCG（S#，C#，G）， $F3\{S\#, C\# \rightarrow G\}$

以上子模式属BCNF且具有无损连接性，**本例函数依赖关系没有损失。**

分解算法

4. 具有无损连接性的4NF分解算法

分解算法: $\rho = \{R_i\}$

WHILE ρ 中存在非4NF关系 DO

 选择一个非4NF模式 $S \in \rho$

 选择 S 的一个违反4NF要求的多值依赖 $X \twoheadrightarrow Y$

 用两个关系模式 $(X \cup Y)$ 和 $(S - Y)$ 取代 ρ 中的 S

ENDWHILE

分解算法

例1: $R(U) = R(A, B, C)$, $D = \{A \twoheadrightarrow B, A \twoheadrightarrow C\}$
指出 $R \notin 4NF$, 分解 R 使属 $4NF$, 具有无损连接性。

(1) $\because R$ 的主键是 A, B, C

$\because R$ 中 $A \twoheadrightarrow B$, A 不是主键 $\therefore R$ 不属于 $4NF$

(2) 对 $A \twoheadrightarrow B$

$R_1(\underline{A}, \underline{B})$

$R_2(\underline{A}, \underline{C}) = U - B,$

$\rho = \{R_1, R_2\} \in 4NF$

分解算法

例2、 $R = (A, B, C)$, $D = \{AB \rightarrow C, C \twoheadrightarrow A\}$

问 $R \in ?$ NF, 将 R 分解到4NF且具有无损连接性。

主键 AB , $R \in \text{BCNF}$

分解 $CA (C, A)$,

$CB (B, C) \in 4\text{NF}$

以上子模式属4NF且具有无损连接性。

本例, 函数依赖关系丢失。

分解算法

规范化理论提供了一套模式分解算法，按这些算法可以做到：

1. 若要求分解具有无损连接性，模式分解一定可以达到4NF (在函数依赖范畴, 一定可达到BCNF)。
2. 若要求分解保持函数依赖，模式分解一定可以达到3NF，但不一定能够达到BCNF。
3. 若要求分解既要保持函数依赖，又具有无损连接性，模式分解一定可以达到3NF，但不一定能够达到BCNF。

小结

- ❖ 规范化理论为数据库设计提供了理论的指南和工具
- ❖ 并不是规范化程度越高，模式就越好
 - 必须结合应用环境和现实世界的具体情况合理地选择数据库模式



小 结

关系数据库中的关系模式，按其属性间满足不同程度的要求，定义了不同级别的范式。

某级范式是满足某级要求的关系模式的集合。

在关系模式中，关系模式的基本要求是满足1NF，这样的关系模式才是合法的。

小 结

- 规范化程度过低的关系不能很好地描述现实世界，可能会存在插入异常、删除异常、修改复杂、数据冗余等问题。
- 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集合，这种过程就叫关系模式的规范化。

小结

消除
不合
适的
函数
依赖

1NF



消除非主属性对候选码的部分函数依赖

2NF



消除非主属性对候选码的传递函数依赖

3NF



消除主属性对候选码的部分或传递函数依赖

BCNF



消除非平凡非函数依赖的多值依赖

4NF

小结

本章在函数依赖与多值依赖的范畴内讨论关系模式的**规范化**。关系模式的规范化过程是通过对关系模式的**分解**来实现的，而数据依赖的**公理系统**是模式分解的理论基础。在关系模式分解中应具有**无损连接性**和**函数依赖保持性**是至关重要的。