

# 数据库系统概论

## An Introduction to Database Systems

---

### 关系数据库规范化理论(二)

The Relational Database Normalization Theory

陆嘉恒

[www.jiahenglu.net](http://www.jiahenglu.net)

中国人民大学信息学院

**NDBC 2008 主旨发言**

# **Review of Claremont Report on Database Research**

**Jiaheng Lu**

**[www.jiahenglu.net](http://www.jiahenglu.net)**

**Renmin University of China**

# Outline

---

- **Five challenges on database research**
  - **Structured and unstructured data**
  - **Cloud data management**
  - **Declarative programming**
  - **Database engine revisiting**
  - **Mobile application**
- **Our research to meet those challenges**

# 数据库的挑战:

## Senior database researcher Meeting

---

- **Senior database researchers have gathered every few years to assess the state of database research and to recommend problems and problem areas deserve additional focus.**
  - **Laguna Beach, Calif. in 1989**
  - **Palo Alto, Calif. (“Lagunita”) in 1990 and 1995**
  - **Cambridge, Mass. in 1996**
  - **Asilomar, Calif. in 1998**
  - **Lowell, Mass . In 2003**

# Claremont Meeting

---

- **About 20 Database researchers**
- **Claremont Resort, Berkeley, CA  
May 29-30, 2008**



# Question?

---

- **Welcome question**

# 复习：容易混淆的概念

---

- 码，主码，候选码，超码？
- 码 = 主码？
- 码 = 候选码？

## 复习：容易混淆的概念(二)

---

- 非主属性，主码，候选码？
- 非主属性 **NOT IN** 主码？
- 非主属性 **NOT IN** 候选码？

## 5.2.2 第二范式 (2NF)

---

### ● 2NF的定义

定义5.8 若关系模式 $R \in 1NF$ ，并且每一个非主属性都完全函数依赖于 $R$ 的码，则 $R \in 2NF$ 。

# 符合第二范式的关系

---

例：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。

- 函数依赖：

假设每一教师只教一门课。每门课由若干教师教，某一学生选定某门课，就确定了一个固定的教师。于是有：

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

## 第二范式（续）

---

- $(S, J)$ 和 $(S, T)$ 都可以作为候选码。
- $STJ \in 2NF$
- $T \rightarrow J$ ,  $J$  部分依赖于 $T$ , 但 $J$ 是主属性集。

# Question?

---

- **Welcome question**

## 5.2 规范化

---

5.2.1 第一范式 (1NF)

5.2.2 第二范式 (2NF)

5.2.3 第三范式 (3NF)

5.2.4 BC范式 (BCNF)

5.2.5 多值依赖与第四范式 (4NF)

5.2.6 规范化

## 5.2.5 多值依赖与第四范式 (4NF)

---

- 例子
- 多值依赖
- 第四范式 (4NF)

# 多值依赖与第四范式（续）

---

## 例子

### 属于BCNF的关系模式:

- 函数依赖: 一个完美的关系模式
- 多值依赖:

例: 设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。

用关系模式Teaching(C, T, B)来表示课程C、教师T和参考书B之间的关系。

## 多值依赖与第四范式（续）

| 表5.1 | 课程 C | 教员 T           | 参考书 B                      |
|------|------|----------------|----------------------------|
|      | 物理   | { 李 勇<br>王 军 } | { 普通物理学<br>光学原理<br>物理习题集 } |
|      | 数学   | { 李 勇<br>张 平 } | { 数学分析<br>微分方程<br>高等代数 }   |
|      | 计算数学 | { 张 平<br>周 峰 } | { 数学分析 }                   |
|      | ⋮    | ⋮              | ⋮                          |

⋮

# 多值依赖与第四范式 (续)

用二维表表示：表5.2 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  |
| ...   | ... | ...   |

## 多值依赖与第四范式（续）

---

- **Teaching** ∈ BCNF: **Teach** 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: 有多少名任课教师, 参考书就要存储多少次。
  - (2) 增加操作复杂: 当某一课程增加一名任课教师时, 该课程有多少本参照书, 就必须插入多少个元组。

# 多值依赖与第四范式（续）

---

例如物理课增加一名教师刘翔，需要插入两个元组：

（物理，刘翔，普通物理学），（物理，刘翔，光学原理）

(3)删除操作复杂：某一门课要去掉一本参考书，该课程有多少名教师，就必须删除多少个元组。

(4)修改操作复杂：某一门课要修改一本参考书，该课程有多少名教师，就必须修改多少个元组。

产生原因：

参考书的取值和教师的取值是彼此独立毫无关系的，都只取决于课程名。

# 一、多值依赖

---

- 定义

定义5.11 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 $Z=U-X-Y$ ，多值依赖 $X \twoheadrightarrow Y$ 成立当且仅当对 $R$ 的任一关系 $r$ ， $r$ 在 $(X, Z)$ 上的每个值对应一组 $Y$ 的值，这组值仅仅决定于 $X$ 值而与 $Z$ 值无关。

例 Teaching (C,T,B)

- 平凡多值依赖和非平凡的多值依赖

若 $X \twoheadrightarrow Y$ ，而 $Z=\varnothing$ ，则称 $X \twoheadrightarrow Y$ 为平凡的多值依赖。

否则称 $X \twoheadrightarrow Y$ 为非平凡的多值依赖。

# 多值依赖（续）

---

## 多值依赖的性质

(1) 多值依赖具有对称性。

若  $X \twoheadrightarrow Y$ ，则  $X \twoheadrightarrow Z$ ，其中  $Z = U - X - Y$

(2) 函数依赖是多值依赖的特殊情况。

若  $X \rightarrow Y$ ，则  $X \twoheadrightarrow Y$ 。

## 二、第四范式 (4NF)

---

### ● 定义

定义5.12 关系模式 $R\langle U, F \rangle \in 1NF$ ，如果对于 $R$ 的每个非平凡多值依赖 $X \twoheadrightarrow Y$  ( $Y \subseteq U \setminus X$ )， $X$ 都含有候选码，则 $R \in 4NF$ 。

4NF就是限制关系模式的属性之间不允许有非平凡且非函数依赖的多值依赖。

## 第四范式（续）

---

- 如果一个关系模式是4NF，则必为BCNF。

例 Teach(C,T,B)中存在非平凡的多值依赖 $C \twoheadrightarrow T$ ，且C不是候选码，因此Teach不属于4NF。这正是它之所以存在数据冗余度大，插入和删除操作复杂等弊病的根源。

## 第四范式（续）

---

- 解决方法

用投影分解法把**Teach**分解为如下两个4NF关系模式：

**CT(C, T)**

**CB(C, B)**

**CT** ∈ 4NF。  $C \twoheadrightarrow T$  是平凡多值依赖，

**CT** 中不存在既非平凡也非函数依赖的多值依赖。

**CB** ∈ 4NF。

## 第四范式（续）

---

分解后**Teach**关系中的几个问题可以得到解决：

- (1) 参考书只需要在**CB**关系中存储一次。
- (2) 当某一课程增加一名任课教师时，只需要在**CT**关系中增加一个元组。
- (3) 某一门课要去掉一本参考书，只需要在**CB**关系中删除一个相应的元组。

# 习题五

---

- 1,2,4,5,9,10,11,12