

数据库系统概论 高级篇

An Introduction to Database System

王珊

中国人民大学信息学院

School of Information,
Renmin University of China

2016

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—为什么要学习关系数据理论
- 6.2 规范化
- 6.3 数据依赖的公理系统（简单介绍）
- 6.4 模式的分解（简单介绍）
- 6.5 小结



第六章 关系数据理论

6.1 问题的提出—为什么要学习关系数据理论

6.2 规范化

6.3 数据依赖的公理系统（简单介绍）

6.4 模式的分解（简单介绍）

6.5 小结



6.1 问题的提出

- 关系数据库的基本概念
- 关系模型
- 关系数据库的标准语言
- 关系数据库逻辑设计

针对一个具体问题，应如何构造一个适合于它的数据模式，即应该构造几个关系，每个关系由哪些属性组成等。

问题——什么是一个好的数据库逻辑设计



一个例子

❖ 学校开发一个学校教务的数据库，涉及的对象有：

学生的学号（Sno）、所在系（Sdept）、系主任姓名（Mname）、课程号（Cno）和成绩（Grade）。

❖ 语义：

1. 一个系有若干学生， 但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。

❖ 存储并管理这些信息

❖ 设计了一个关系模式

STUDENT (Sno ,Sdept, Mname,Cno,Grade)



一个例子

表6.1 Student表

Sno	Sdept	Mname	Cno	Grade
S1	计算机系	张明	C1	95
S2	计算机系	张明	C1	90
S3	计算机系	张明	C1	88
S4	计算机系	张明	C1	70
S5	计算机系	张明	C1	78
:	:	:	:	:
:	:	:	:	:



关系模式存在的问题

关系模式STUDENT (Sno , Sdept, Mname, Cno, Grade) 中存在的问题:

1. 数据冗余度太大，浪费存储空间

如：系主任的姓名重复出现，重复次数与该系所有学生的所有课程成绩出现次数相同。

2. 更新异常 (Update Anomalies)

数据冗余，更新数据时，维护数据完整性代价大

如果某系更换系主任，系统必须修改与该系学生有关的每一个元组。

Sno	Sdept	Mname	Cno	Grade
S1	计算机系	张明	C1	95
S2	计算机系	张明	C1	90
S3	计算机系	张明	C1	88
S4	计算机系	张明	C1	70
S5	计算机系	张明	C1	78



关系模式存在的问题

3. 插入异常 (Insertion Anomalies)，该插入的数据插不进去

如果新成立一个软件工程系，还没有招生，我们就无法把这个系及其系主任的信息存入数据库。

4. 删除异常 (Deletion Anomalies)，不该删除的数据也删去了

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。

Sno	Sdept	Mname	Cno	Grade
S1	计算机系	张明	C1	95
S2	计算机系	张明	C1	90
S3	计算机系	张明	C1	88
S4	计算机系	张明	C1	70
S5	计算机系	张明	C1	78

NULL	软件工程系	严广	null	null
------	-------	----	------	------



STUDENT (Sno , Sdept,
Mname, Cno, Grade)
不是一个好的关系模式。

关系模式存在的问题

❖ 什么是一个好的模式。

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。

❖ 问题的原因

由于模式中的某些数据依赖引起的。

解决方法：把这个单一模式分成3个关系模式：

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

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

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

这3个模式不会发生插入异常、删除异常毛病；数据冗余得到控制。

再把观察、经验上升为理论：

用规范化理论改造关系模式，消除其中不合适的数据依赖。



6.1 问题的提出

1. 问题——什么是一个好的数据库逻辑设计
2. 什么是数据依赖
3. 关系模式的简化表示



2. 什么是数据依赖：一个例子

❖ STUDENT (Sno , Sdept, Mname, Cno, Grade)

❖ 该关系模式的属性集合记为U:

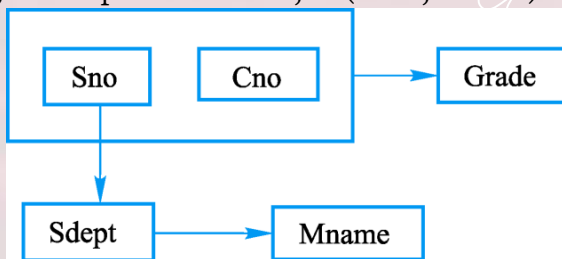
$$U = \{ \text{Sno}, \text{Sdept}, \text{Mname}, \text{Cno}, \text{Grade} \}$$

- 数学中的函数 $y=f(x)$ ，自变量 x 确定之后，相应的函数值 y 也就唯一地确定了。
- $\text{Sdept}=f(\text{Sno})$, Sno函数确定Sdept, 记为 $\text{Sno} \rightarrow \text{Sdept}$,
- $\text{Mname}=f(\text{Sdept})$, Sdept函数确定Mname, 记为 $\text{Sdept} \rightarrow \text{Mname}$
- $\text{Grade}=f((\text{Sno}, \text{Cno}))$, (Sno, Cno)函数确定Grade, 记为 $(\text{Sno}, \text{Cno}) \rightarrow \text{Grade}$

❖ 属性组U上的函数依赖集合，记为F:

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

用一个图表示函数依赖:



什么是数据依赖

1. 完整性约束的一种表现形式

- 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）
即通过属性间值的相等与否来描述
- 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

- 是通过一个关系中属性间值的相等与否体现出来的数据间的相互关系
- 是现实世界属性间相互联系的抽象
- 是数据内在的性质
- 是语义的体现



什么是数据依赖

3. 数据依赖的主要类型

- 函数依赖 (Functional Dependency, 简记为FD)
- 多值依赖 (Multivalued Dependency, 简记为MVD)
- 连接依赖
-

4. 数据依赖对关系模式的影响

- 不合适的数据依赖, 造成插入异常、删除异常、更新异常和数据冗余问题



6.1 问题的提出

- 1.问题——什么是一个好的数据库逻辑设计
- 2.什么是数据依赖
- 3.关系模式的简化表示



3. 关系模式的简化表示

1. 关系模式的形式化定义

(2.1.2关系模式 P.41讲解过)

$R(U, D, DOM, F)$

- R: 关系名, 是符号化的元组语义
- U: 该关系的属性集合
- 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

2. 关系模式的简化表示

$R\langle U, F \rangle$

将关系模式简化为一个三元组, 影响数据库模式设计的主要是 U 和 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。



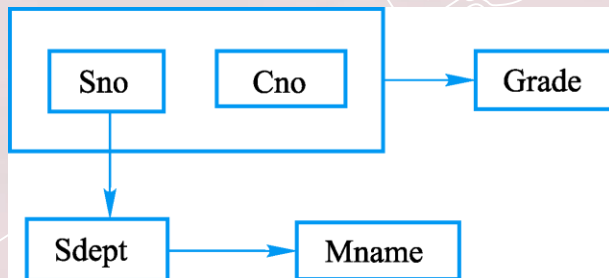
再看例子

关系模式 $STUDENT\langle U, F \rangle$

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

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

$STUDENT (Sno, Sdept, Mname, Cno, Grade, Sno \rightarrow Sdept, Sdept \rightarrow Mname, (Sno, Cno) \rightarrow Grade)$



关系模式 $STUDENT\langle U, F \rangle$ 存在诸多问题。

如何解决关系模式中存在的问题？

规范化理论—找出关系模式中不合适的数据依赖，消除它们，可以在不同程度上解决插入异常、删除异常、更新异常和数据冗余问题。



第六章 关系数据理论

6.1 问题的提出—为什么要学习关系数据理论

6.2 规范化

6.3 数据依赖的公理系统（简单介绍）

6.4 模式的分解（简单介绍）

6.5 小结



思考与讨论

在前面讲解的关系模式STUDENT (Sno , Sdept, Mname, Cno, Grade) 中，如果新成立一个软件工程系，还没有招生，就无法把这个系及其系主任的信息存入数据库。请问为什么？

Sno	Sdept	Mname	Cno	Grade
S1	计算机系	张明	C1	95
S2	计算机系	张明	C1	90
S3	计算机系	张明	C1	88
S4	计算机系	张明	C1	70
S5	计算机系	张明	C1	78

NULL	软件工程系	严广	null	null
------	-------	----	------	------



第六章 关系数据理论

