



南京大学软件学院
NANJING UNIVERSITY · SOFTWARE INSTITUTE



数据管理基础

ch32 关系模式及范式

Software Institute
Nanjing University
Bei Jia

关系模式及第一范式 (1NF)

- 关系模式由五部分组成，是一个五元组：

$$R(U, D, \text{DOM}, F)$$

- 关系名 R 是符号化的元组语义
 - U 为一组属性
 - D 为属性组 U 中的属性所来自的域
 - DOM 为属性到域的映射
 - F 为属性组 U 上的一组数据依赖
- 由于 D 、 DOM 与模式设计关系不大，因此可以把关系模式看作一个三元组： $R\langle U, F \rangle$
 - 当且仅当 U 上的一个关系 r 满足 F 时， r 称为关系模式 $R\langle U, F \rangle$ 的一个关系
 - 作为二维表，关系要符合一个最基本的条件：每个分量必须是不可分开的数据项。满足了这个条件的关系模式就属于第一范式 (1NF)



数据依赖

- 数据依赖

- 是一个关系内部属性与属性之间的一种约束关系
 - ◆ 通过属性间值的相等与否体现出来的数据间相关联系
- 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

- 数据依赖的主要类型

- 函数依赖 (Functional Dependency, 简记为FD)
- 多值依赖 (Multi-Valued Dependency, 简记为MVD)



函数依赖

- 函数依赖普遍存在于现实生活中

- 描述一个学生关系，可以有学号、姓名、系名等属性。

- ◆ 一个学号只对应一个学生，一个学生只在一个系中学习

- ◆ “学号”值确定后，学生的姓名及所在系的值就被唯一确定。

- ◆ $Sname=f(Sno)$, $Sdept=f(Sno)$

- 即Sno函数决定Sname, Sno函数决定Sdept

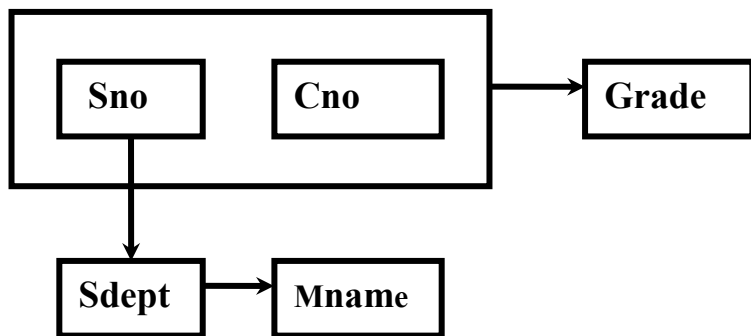
- 记作 $Sno \rightarrow Sname$, $Sno \rightarrow Sdept$



1NF

- [例6.1] 建立一个描述学校教务的数据库。涉及的对象包括：

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



- 假设学校教务的数据库模式用一个单一的关系模式 Student 来表示，则该关系模式的属性集合为：
 - $U = \{Sno, Sdept, Mname, Cno, Grade\}$
- 现实世界的已知事实（语义）：
 - 一个系有若干学生， 但一个学生只属于一个系；
 - 一个系只有一名（正职）负责人；
 - 一个学生可以选修多门课程， 每门课程有若干学生选修；
 - 每个学生学习每一门课程有一个成绩。
- 由此可得到属性组U上的一组函数依赖F：
 - $F = \{Sno \rightarrow Sdept, Sdept \rightarrow Mname, (Sno, Cno) \rightarrow Grade\}$

1NF的问题 1

- 关系模式Student<U, F>中存在的问题:

- 数据冗余

- ◆ 浪费大量的存储空间

- 每一个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。

- 修改复杂, 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。

- 某系更换系主任后, 必须修改与该系学生有关的每一个元组。

- 插入异常 (Insertion Anomalies)

- ◆ 如果一个系刚成立, 尚无学生, 则无法把这个系及其系主任的信息存入数据库。

- 删除异常 (Deletion Anomalies)

- ◆ 如果某个系的学生全部毕业了, 则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。



1NF的问题 2

- 结论
 - Student关系模式不是一个好的模式。
 - 一个“好”的模式应当不会发生插入异常、删除异常和更新异常，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
 - 用规范化理论改造关系模式来消除其中不合适的数据依赖
- 把这个单一的模式分成三个关系模式：
 - $S(Sno, Sdept, Sno \rightarrow Sdept);$
 - $SC(Sno, Cno, Grade, (Sno, Cno) \rightarrow Grade);$
 - $DEPT(Sdept, Mname, Sdept \rightarrow Mname);$
- 这三个模式都不会发生插入异常、删除异常的问题，数据的冗余也得到了控制。



范式 1

- 范式是符合某一种级别的关系模式的集合。
- 关系数据库中的关系必须满足一定的要求。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)



范式 2

- 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（schema decomposition）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（normalization）。

