

7.5 规范化

主要考点

- 1、1NF（第一范式）
- 2、2NF（第二范式）
- 3、3NF（第三范式）
- 4、BCNF（巴克斯范式）
- 5、4NF（第四范式）

1NF（第一范式）

- 定义：若关系模式R的每一个分量是不可再分的数据项，则关系模式R属于第一范式。记为 $R \in 1NF$ 。

例：学生(学号, 姓名, 学院编号, 学院名称, 课程号, 成绩)

$F = \{\text{学号} \rightarrow \text{姓名}, \text{学号} \rightarrow \text{学院编号}, \text{学院编号} \rightarrow \text{学院名称}, (\text{学号}, \text{课程号}) \rightarrow \text{成绩}\}$

- 存在的问题：

- (1) 数据冗余。
- (2) 更新异常(修改操作后数据不一致)。
- (3) 插入异常。
- (4) 删除异常。

学生

学号	姓名	学院编号	学院名称	课程号	成绩
1001	王芳	D20	经济学院	C001	76
1001	王芳	D20	经济学院	C002	85
1001	王芳	D20	经济学院	C003	63
1002	李明	D12	计算机学院	C001	92
1002	李明	D12	计算机学院	C002	89
1003	赵晗	D12	计算机学院	C002	78
1003	赵晗	D12	计算机学院	C003	86
1004	刘晓	D25	管理学院	C002	90

2NF（第二范式）

- 定义：若关系模式 $R \in 1NF$ ，且每一个非主属性完全依赖于码，则关系模式 $R \in 2NF$ 。
- 换句话说：当1NF消除了非主属性对码的部分函数依赖，则称为2NF。

例：学生(学号, 姓名, 学院编号, 学院名称, 课程号, 成绩)

$F = \{\text{学号} \rightarrow \text{姓名}, \text{学号} \rightarrow \text{学院编号}, \text{学院编号} \rightarrow \text{学院名称}, (\text{学号}, \text{课程号}) \rightarrow \text{成绩}\}$

将学生关系分解为：

- 学生₁(学号, 姓名, 学院编号, 学院名称) $\in 2NF$
- 学生₂(学号, 课程号, 成绩) $\in 2NF$

学生₁

学号	姓名	学院编号	学院名称
1001	王芳	D20	经济学院
1002	李明	D12	计算机学院
1003	赵晗	D12	计算机学院
1004	刘晓	D25	管理学院

学生₂

学号	课程号	成绩
1001	C001	76
1001	C002	85
1001	C003	63
1002	C001	92
1002	C002	89
1003	C002	78
1003	C003	86
1004	C002	90

3NF（第三范式）

- 定义：若关系模式 $R(U, F)$ 中不存在这样的码 X ，属性组 Y 及非主属性 Z （ $Z \notin Y$ ）使得 $X \rightarrow Y$, $(Y \twoheadrightarrow X) \ Y \rightarrow Z$ 成立，则关系模式 $R \in 3NF$ 。
- 即：当2NF消除了非主属性对码的传递函数依赖，则称为3NF。

例：学生₁ (学号, 姓名, 学院编号, 学院名称) $\in 2NF$, 但 $\notin 3NF$ 。

将学生₁分解为：

- 学生₁₁ (学号, 姓名, 学院编号) $\in 3NF$
- 学生₁₂ (学院编号, 学院名称) $\in 3NF$

学生₁₁

学号	姓名	学院编号
1001	王芳	D20
1002	李明	D12
1003	赵晗	D12
1004	刘晓	D25

学生₁₂

学院编号	学院名称
D20	经济学院
D12	计算机学院
D25	管理学院

BCNF（巴克斯范式）

- 定义：关系模式 $R \in 1NF$ ，若 $X \rightarrow Y$ 且 $Y \not\subseteq X$ 时， X 必含有码，则关系模式 $R \in BCNF$ 。
- 也就是说，当3NF消除了主属性对码的部分函数依赖和传递函数依赖，则称为BCNF。
- 结论：一个满足BCNF的关系模式，应有如下性质：
 - (1) 所有非主属性对每一个码都是完全函数依赖；
 - (2) 所有主属性对每一个不包含它的码，也是完全函数依赖；
 - (3) 没有任何属性完全函数依赖于非码的任何一组属性。

例：R(Pno, Pname, Mname)的属性表示零件号、零件名和厂商名，如果约定，每种零件号只有一个零件名，但不同的零件号可以有相同的零件名；每种零件可以有多个厂商生产，但每家厂商生产的零件应有不同的零件名。函数依赖集如下：

$(Pname, Mname) \rightarrow Pno, Pno \rightarrow Pname$

候选码为(Pname, Mname)或(Pno, Mname)

分解为：R1(Pno, Pname) $\in BCNF$

R2(Pno, Mname) $\in BCNF$

4NF（第四范式）

- 定义：关系模式 $R \in 1NF$ ，若对于 R 的每个非平凡多值依赖 $X \twoheadrightarrow Y$ 且 $Y \not\subseteq X$ 时， X 必含有码，则关系模式 $R(U, F) \in 4NF$ 。
- 4NF是限制关系模式的属性间不允许有非平凡且非函数依赖的多值依赖。
- 注意：如果只考虑函数依赖，关系模式最高的规范化程度是BCNF，如果考虑多值依赖，关系模式最高的规范化程度是4NF。

例：学生(学号, 选修课程, 兴趣爱好)

- 学生₁(学号, 选修课程) $\in 4NF$
- 学生₂(学号, 兴趣爱好) $\in 4NF$

书目(ISBN, 书名, 作者, 排名, 出版社)

- 书目₁(ISBN, 作者, 排名) $\in 4NF$
- 书目₂(ISBN, 书名, 出版社) $\in BCNF$

学生

学号	选修课程	兴趣爱好
1001	C001	唱歌
1001	C002	跳舞
1002	C003	羽毛球
1002	C001	跆拳道
1003	C002	画画
1003	C003	跳舞

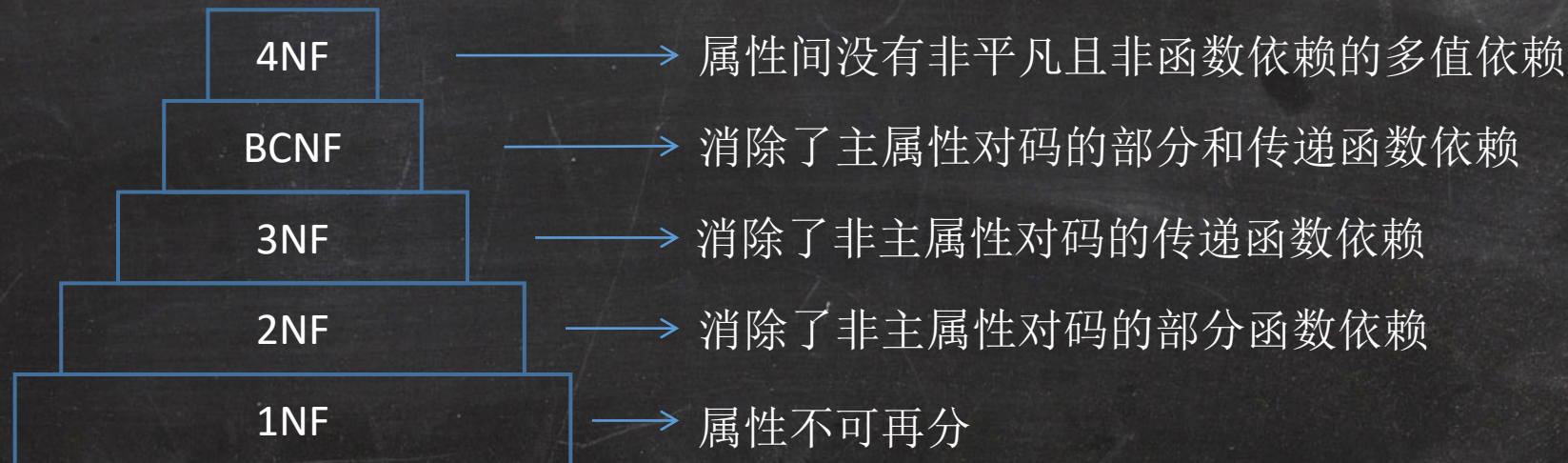
书目₁

ISBN	作者	排名
1039	李晓	1
1039	王文	2
1039	赵军	3
1040	刘莉	1
1040	艾青	2

总结

$1NF \supset 2NF \supset 3NF \supset BCNF \supset 4NF \supset 5NF$

通过分解，可以将一个低一级范式的关系模式转换成若干个高一级范式的关系模式，这种过程叫做**规范化**。



例：设有关系模式R（课程，教师，学生，成绩，时间，教室），其中函数依赖集F如下：

$F = \{ \text{课程} \twoheadrightarrow \text{教师}, (\text{学生}, \text{课程}) \rightarrow \text{成绩}, (\text{时间}, \text{教室}) \rightarrow \text{课程}, (\text{时间}, \text{教师}) \rightarrow \text{教室}, (\text{时间}, \text{学生}) \rightarrow \text{教室} \}$

关系模式R的一个主键是(D)，R规范化程序最高达到(B)。若将关系模式R分解为3个关系模式R1（课程，教师）、R2（学生，课程，成绩）、R3（学生，时间，教室，课程），其中R2的规范化程度最高达到(C)。

- (1) A. (学生，课程) B. (时间，教室)
C. (时间，教师) D. (时间，学生)
- (2) A. 1NF B. 2NF C. 3NF D. BCNF
- (3) A. 2NF B. 3NF C. BCNF D. 4NF