

第六章 关系数据理论

范 式

一、关系模式相关定义

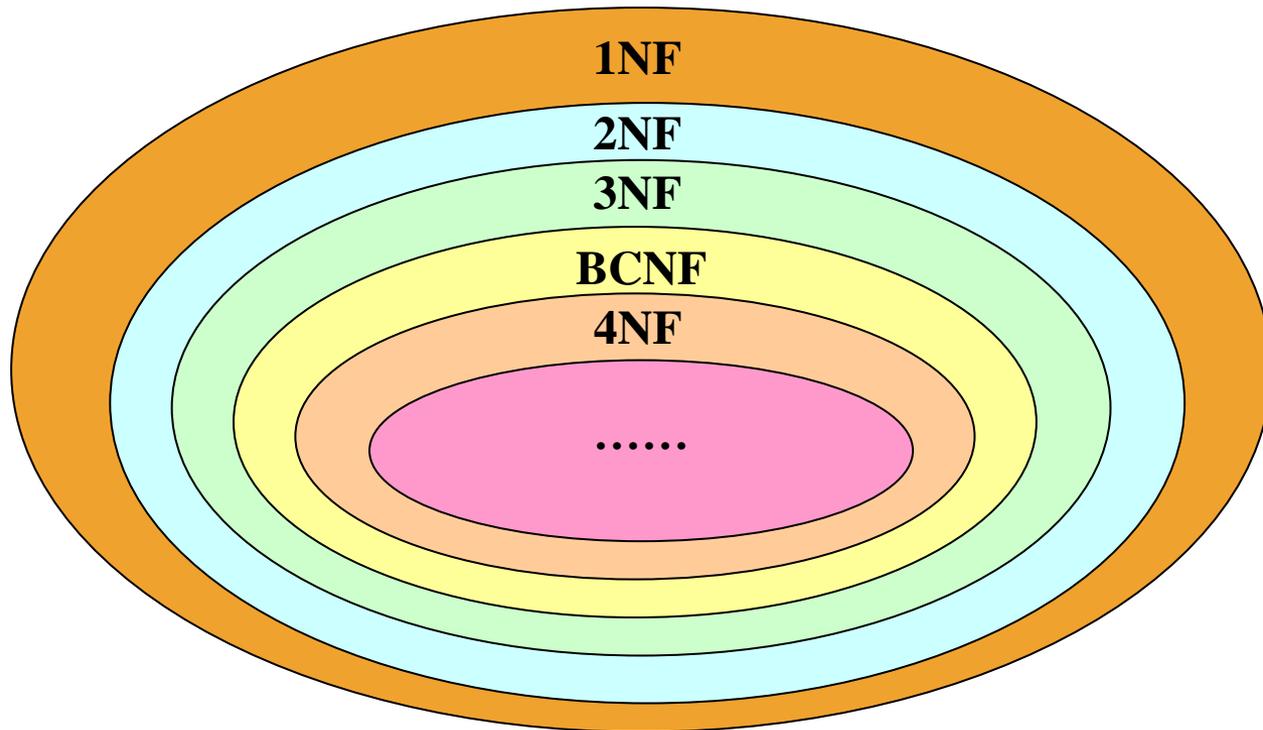
- **定义6.4:** 设K为R (U,F) 中的属性或属性组合。若 $K \xrightarrow{F} U$, 则K称为R的**候选码**; 若候选码多于一个, 则选定其中的一个做为**主码**;
- 主属性与非主属性
 - 包含在任何一个候选码中的属性, 称为**主属性**;
 - 不包含在任何码中的属性称为**非主属性**或**非码属性**;
- 最简单情况: 一个属性是码;
- 最极端情况: 整个属性组是码, 也称**全码**;

- 定义6.5：关系模式 R 中属性或属性组 X 并非 R 的码，但 X 是另一个关系模式的码，则称 X 是 R 的**外部码**，也称**外码**；
- 如在 SC (Sno, Cno, Grade) 中，Sno 不是码，但 Sno 是关系模式 S (Sno, Sdept, Sage) 的码，则 Sno 是关系模式 SC 的外部码；
- 主码与外部码一起提供了表示关系间联系的手段。

二、关系模式规范化

- 关系模式的好与坏，用什么标准衡量？这个标准就是模式的**范式**（简记为**NF**）；
- 范式是符合某一种级别的关系模式的集合；
- 范式的种类与数据依赖有着直接的联系，基于FD的范式有**1NF**、**2NF**、**3NF**、**BCNF**等多种。
- 某一关系模式R为第n范式，可简记为 **$R \in nNF$** 。

- 各范式之间的关系



- 如果一个关系满足某个范式要求, 则它也会满足**较其级别低**的所有范式的要求。
- 一个低一级范式的关系模式, 通过**模式分解**（投影运算）可以转化为若干个高一级范式的关系模式的集合, 这个过程叫做**规范化**;
- 用几个简单的关系去取代原来结构复杂的的关系的过程叫做**关系规范化**。
- 规范化理论是围绕着范式建立的。

关系规范化

- **目的**: 使结构合理, 使数据冗余尽量小, 减少插入, 删除和更新异常;
- **方法**: 将关系模式投影分解成多个关系模式, 但关系模式的分解不是唯一的;
- **要求**: 分解后的关系模式集合应当与原关系模式“等价”, 既具有无损连接性, 又保持函数依赖特性。

1) 第一范式1NF

- 当且仅当一个关系R中，每一个分量都是不可再分的数据项时， $R \in 1NF$ ；
- 要求属性是原子的；
- 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。

2) 第二范式2NF

- 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 $R \in 2NF$ ；
- 如果关系模式 $R \in 1NF$ ，并且R的码是单个属性，那么 $R \in 2NF$ 。

例1：判断关系模式S-L-C (Sno, Sdept, Sloc, Cno, Grade) 是否满足2NF, 其中 Sloc 为学生住处, 假设每个系的学生住在同一个地方;

- ✓ 第一步：判断S-L-C是否为第一范式;
- ✓ 第二步：确定候选码 (Sno, Cno) ;
- ✓ 第三步：确定所有函数依赖有： Sno \rightarrow Sdept, Sno \rightarrow Sloc,
(Sno , Cno) \xrightarrow{P} Sdept, (Sno , Cno) \xrightarrow{P} Sloc ,
(Sno , Cno) \xrightarrow{F} Grade;

得出结论：和2NF定义相违背, 故S-L-C不是2NF;

例2. 判断关系模式 $sc(sno, sname, cno, grade)$ 是否满足2NF;

✓ 第一步: 判断 sc 是否为第一范式;

✓ 第二步: 确定候选码 (sno, cno) ;

✓ 第三步: 确定所有函数依赖有: $sno \rightarrow sname$,

$$(sno, cno) \xrightarrow{P} sname , \quad (sno, cno) \xrightarrow{F} grade ;$$

得出结论: 和2NF定义相违背, 故 sc 不是2NF;

一个关系模式不是 2NF，就会产生以下问题（S-L-C）：

- (1) **插入异常**：假设 Sno=95102，Sdept=IS，Sloc=N 的学生还未选课，因课程号是主属性，因此该学生的信息无法插入SLC。
- (2) **删除异常**：假定某个学生本来只选修了3号课程这一门课。现在因其他原因，3号课程也不选修了。因课程号是主属性，此操作将导致该学生信息的整个元组都要删除。
- (3) **数据冗余度大**：如果一个学生选修了10门课程，那么他的Sdept和Sloc值就要重复存储了10次。
- (4) **修改复杂**：例如学生转系，在修改此学生元组的Sdept值的同时，还可能需修改住处（Sloc）。如果这个学生选修了K门课，则必须无遗漏地修改K个元组中全部Sdept、Sloc信息。

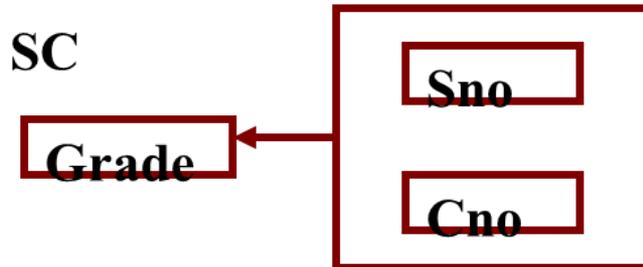
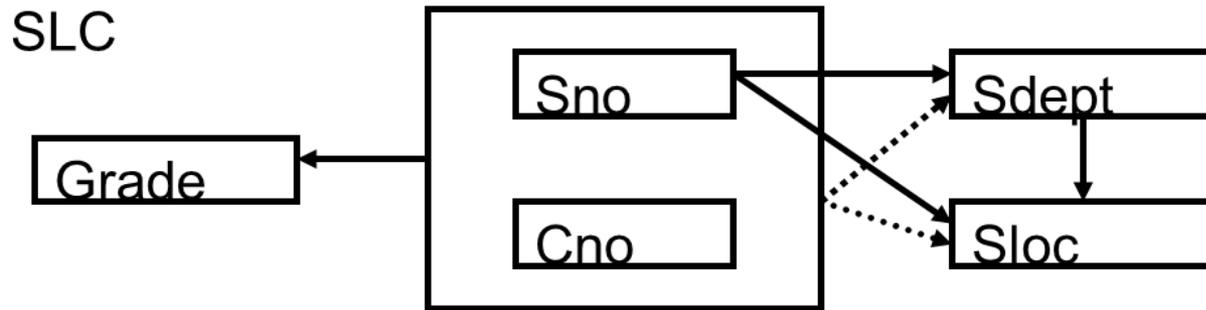
- 原因：非主属性Sdept、 Sloc**部分函数依赖**于码。
- 解决方法：S-L-C分解为两个关系模式，以消除这些部分函数依赖；

SC (Sno, Cno, Grade)

S-L (Sno, Sdept, Sloc)

分解成2NF模式集的方法：投影分解法

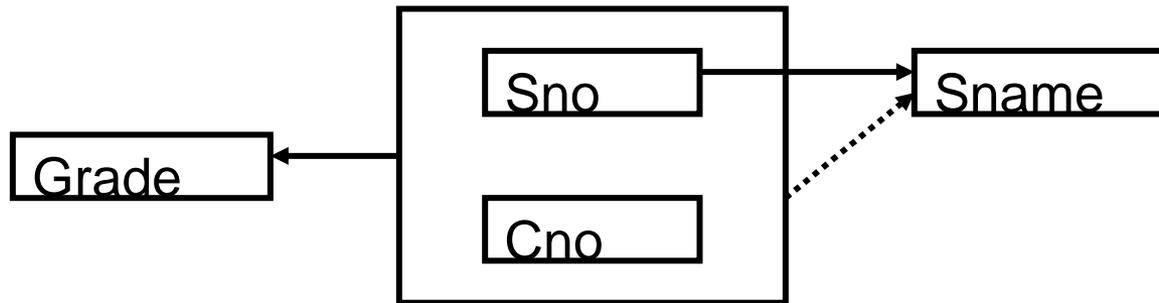
S-L-C函数依赖图为：



- SC (Sno, Cno, Grade) 和 S-L (Sno, Sdept, Sloc)

例2. 分解关系模式 $sc (sno, sname, cno, grade)$ 为2NF;

- 函数依赖图为:



- 可分解为:

模式S ($sno, sname$)

模式SC ($sno, cno, grade$)

- 采用投影分解法将一个1NF的关系分解为多个2NF的关系，可以在一定程度上**减轻**原1NF关系中存在的插入异常、删除异常、数据冗余度大、修改复杂等问题。
- 将一个1NF关系分解为多个2NF的关系，**并不能完全消除**关系模式中的各种异常情况和数据冗余。如S-L (Sno, Sdept, Sloc) 。

3) 第三范式3NF

- **定义6.7**：如果关系模式R是1NF，且每个**非主属性都不传递依赖于R的候选码**，那么称R是第三范式（3NF）的模式。
- 如果 $R \in 3NF$ ，则R也是2NF。
- 若 $R \in 3NF$ ，则每一个**非主属性既不部分依赖于码也不传递依赖于码**；

- **例1：判断2NF关系模式SC (sno, cno, grade) 及S-L (sno, sdept, sloc) 是否满足3NF；**

模式SC的函数依赖有： $(sno, cno) \rightarrow grade$

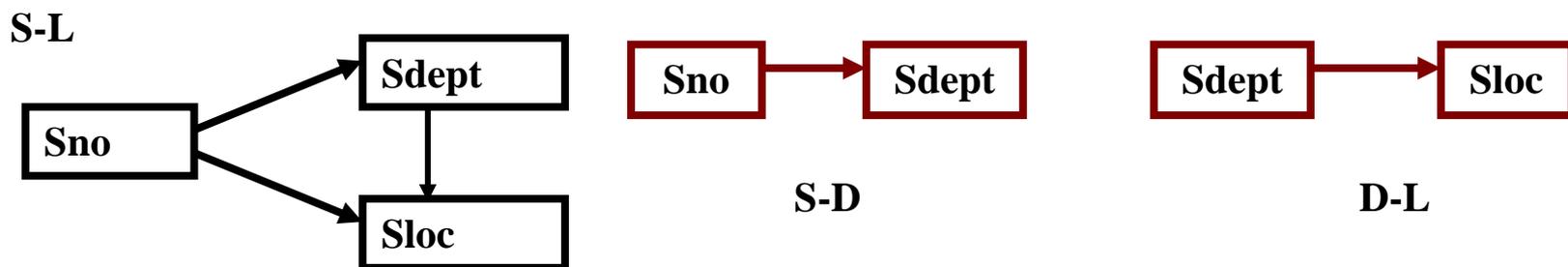
模式S-L的函数依赖有：

$sno \rightarrow sdept, sdept \not\rightarrow Sno, sdept \rightarrow sloc$

故：模式SC为3NF，模式S-L依然为2NF；

- 解决方法:

采用投影分解法, 把S-L分解为两个关系模式, 以消除传递函数依赖:



S-D (Sno, Sdept), D-L (Sdept, Sloc)

- 分解后的关系模式S-D与D-L中不再存在传递依赖

- S-D (Sno, Sdept) \in 3NF, D-L (Sdept, Sloc) \in 3NF

- **例2：判断2NF关系模式S (sno, sname) 及SC (Sno, Cno, Grade) 是否满足3NF；**

模式sc的函数依赖有： $sno \rightarrow sname$

模式sc的函数依赖有： $(sno, cno) \rightarrow grade$

故：模式s和sc皆为3NF；

- 采用投影分解法将一个2NF的关系分解为多个3NF的关系，可以在一定程度上**解决原2NF关系中**存在的插入异常、删除异常、数据冗余度大、修改复杂等问题。
- 将一个2NF关系分解为多个3NF的关系后，仍然不能完全消除关系模式中的各种异常情况和数据冗余。可能表现在**主属性对码的部分依赖和传递依赖**。

4) BCNF

- **定义6.8** 设关系模式 $R(U, F) \in 1NF$ ，如果对于R的**每个函数依赖 $X \rightarrow Y$** ，若Y不属于X，则X必含有候选码，那么 $R \in BCNF$ 。
- 若 $R \in BCNF$
 - **每一个决定属性集（因素）都包含（候选）码；**
 - R中的所有属性（**主，非主属性**）都完全函数依赖于码；
 - $R \in 3NF$;

BCNF的关系模式所具有的性质

1. 所有**非主属性**都**完全函数依赖于**每个**候选码**
2. 所有**主属性**都**完全函数依赖于**每个**不包含它的候选码**
3. **没有任何属性完全函数依赖于非码的任何一组属性**

- 例1: 判断3NF模式S-D (Sno, Sdept), D-L (Sdept, Sloc) 及 SC (Sno, Cno, Grade) 是否满足BCNF;

判断

- ✓ 每一个决定属性集 (因素) 都包含 (候选) 码;
- ✓ 主属性对码是否存在部分依赖和传递依赖;

对于s-d模式, 满足, 故属于BCNF;

对于d-l模式, 满足, 故属于BCNF;

对于sc模式, 满足, 故属于BCNF;

- **例2：判断3NF模式S (sno, sname) 及SC (Sno, Cno, Grade) 是否满足BCNF；**

判断

- ✓ 每一个决定属性集（因素）都包含（候选）码；
- ✓ 主属性对码是否存在部分依赖和传递依赖；

对于s模式，满足，故属于BCNF；

对于sc模式，满足，故属于BCNF；

例3：判断关系模式C (Cno, Cname, Peno)

- $C \in 3NF$
- $C \in BCNF$

例4：判断关系模式S (Sno, Sname, Sdept, Sage)

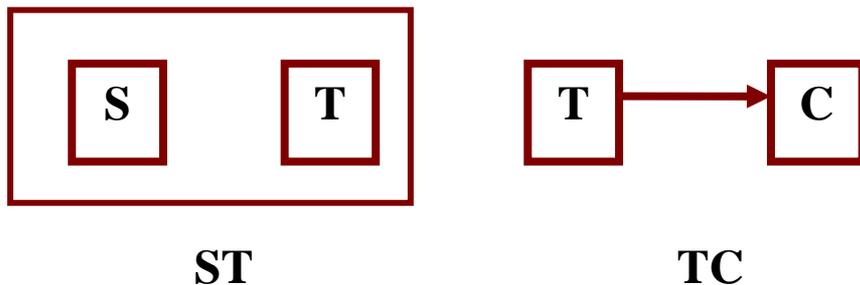
- 假定Sname唯一
- 假定S有两个码Sno, Sname
- $S \in 3NF$
- $S \in BCNF$

- 例5: 关系模式SPJ (S, P, J) , S: 学生, J: 课程, P: 名次, 每一个学生选修每门课程的成绩有一定的名次, 每门课程中每一名次只有一个学生 (无并列名次) ;
 - 函数依赖: $(S, J) \rightarrow P$; $(J, P) \rightarrow S$
 - (S, J) 与 (J, P) 都可以作为候选码, 属性相交, 属性间无部分依赖或传递依赖
 - $SJP \in 3NF$,
 - $SJP \in BCNF$

- **例6：在关系模式STC (S, T, C) 中，S表示学生，T表示教师，C表示课程。每一教师只教一门课，每门课有若干教师，某学生选修某门课就对应一个老师。**
- 函数依赖有： $(S, C) \rightarrow T$ ， $(S, T) \rightarrow C$ ， $T \rightarrow C$
- 候选码： (S, T) ， (S, C)
- 无非主属性对码的部分依赖和传递依赖，但是存在主属性部分依赖于码的情况；
- 故不属于BCNF；

- 解决方法：将STC分解为二个关系模式：

$ST(S, T) \in BCNF, TC(T, C) \in BCNF$



没有**任何属性**对码的部分函数依赖和传递函数依赖

- 一个关系模式达到BCNF，说明在函数依赖的范畴内，已实现了彻底分离，可消除“异常”，但在实际应用中，并不一定要求全部模式都达到BCNF。

3NF与BCNF的关系

- 如果关系模式 $R \in \text{BCNF}$ ，必定有 $R \in \text{3NF}$
- 如果 $R \in \text{3NF}$ ，且 R 只有一个候选码，则 R 必属于BCNF。

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



练习题

1、设有关系模式R（学号Sno，课程号Cno，成绩G，任课教师TN，教师专长TS），基于R的函数依赖集 $F=\{ (Sno, Cno) \rightarrow G, Cno \rightarrow TN, TN \rightarrow TS \}$ ，判断R为第几范式？

判断步骤：

- ① 是否为1NF；
- ② 候选码有（Sno，Cno），非主属性为（G，TN和TS）
- ③ 判断函数依赖集中非主属性是否完全函数依赖于候选码；非主属性G对键是完全依赖： $(Sno, Cno) \rightarrow G$ 。但非主属性TN,TS对键是部分依赖(他们仅依赖于键的真子集Cno)。
- ④ 得出结论：属于1NF；

2、设有关系模式R (sno, sname, sdept) , 函数依赖集F={sno→sname , sno→sdept}, 判断R为第几范式?

判断步骤:

- ① 是否为1NF;
- ② 候选码为 (sno) , 非主属性为 (sname和 sdept)
- ③ 判断函数依赖集中非主属性是否完全函数依赖于候选码; 非主属性sname、sdept对码是完全依赖, 属于2NF;
- ④ 非主属性 (sname和sdept) 是否存在传递依赖候选码→属于3NF;
- ⑤ 主属性sno 是否存在对码的部分和传递依赖 →属于BCNF;
- ⑥ 得出结论: 属于BCNF

3、设有关系模式R（课程C，教师T，参考书B），学校中某一门课程由多个教师讲授，使用相同的一套参考书；教师可讲授多门课程，每种参考书也可以供多门课程使用。判断R为第几范式？

判断步骤：

- ① 是否为1NF；
- ② 候选码为（C,T,B），为全码；
- ③ 属于3NF；
- ④ 主属性是否存在对码的部分和传递依赖 →属于BCNF；
- ⑤ 得出结论：属于BCNF

- 以上我们完全是在**函数依赖**的范畴内讨论问题。属于BCNF的关系模式是否就很完美了呢？
- 一个关系模式达到BCNF，说明在函数依赖的范畴内，已实现了彻底分离，可消除“异常”；
- 但从练习3可以看出，在关系 R 中数据冗余和操作异常都十分明显。究其原因，是关系 R 中存在着一种称之为**多值依赖**的数据依赖。
(如课程：教师=m：n 等)