第四章 关系数据理论

本章导读

教学内容

- 函数依赖
- 规范化设计
- 数据依赖的公理系统

要求掌握

- 1、函数依赖和多值依赖
- 2、各级范式的性质
- 3、低级范式向高级范式的转化

教学重点及难点

- 函数依赖、范式

本章内容

第一节　问题的提出

第二节　函数依赖

第三节　范式

第四节　数据依赖的公理系统

第一节　关系模式设计问题的提出

1、如何设计一个合理的关系数据库模式?

例：设计一个关系数据模式以存放学生各门课考试成绩。

有若干答案:

1) STUDENT关系：SNO, SNAME
   COURSE关系: CNO, CNAME
   GRADE关系: SNO, CNO, GRADE
2) GRADE关系: SNO, SNAME, CNO, GRADE
   COURSE关系: CNO, CNAME
3) GRADE关系: SNO, CNO, CNAME, GRADE
   STUDENT关系：SNO, SNAME
4) GRADE关系: SNO, SNAME, CNO, CNAME, GRADE

泛关系模式

关系模式设计问题的提出

我们称前面的模式为**泛关系模式**，它把现实问题的所有属性组成一个关系模式。

泛关系:

sno	sname	cno	cname	grade
s1	李立	c1	DB	78
s1	李立	c3	OS	87
s2	丁惠	c1	DB	90
s2	丁惠	c2	DS	83
s2	丁惠	c3	OS	77

泛关系模式中存在的问题

sno	sname	cno	cname	grade
s1	李立	c1	DB	78
s1	李立	c3	OS	87
s2	丁惠	c1	DB	90
s2	丁惠	c2	DS	83
s2	丁惠	c3	OS	77

(1)数据冗余　(2)更新异常

(3)插入异常　(4)删除异常

关系模式设计问题的提出

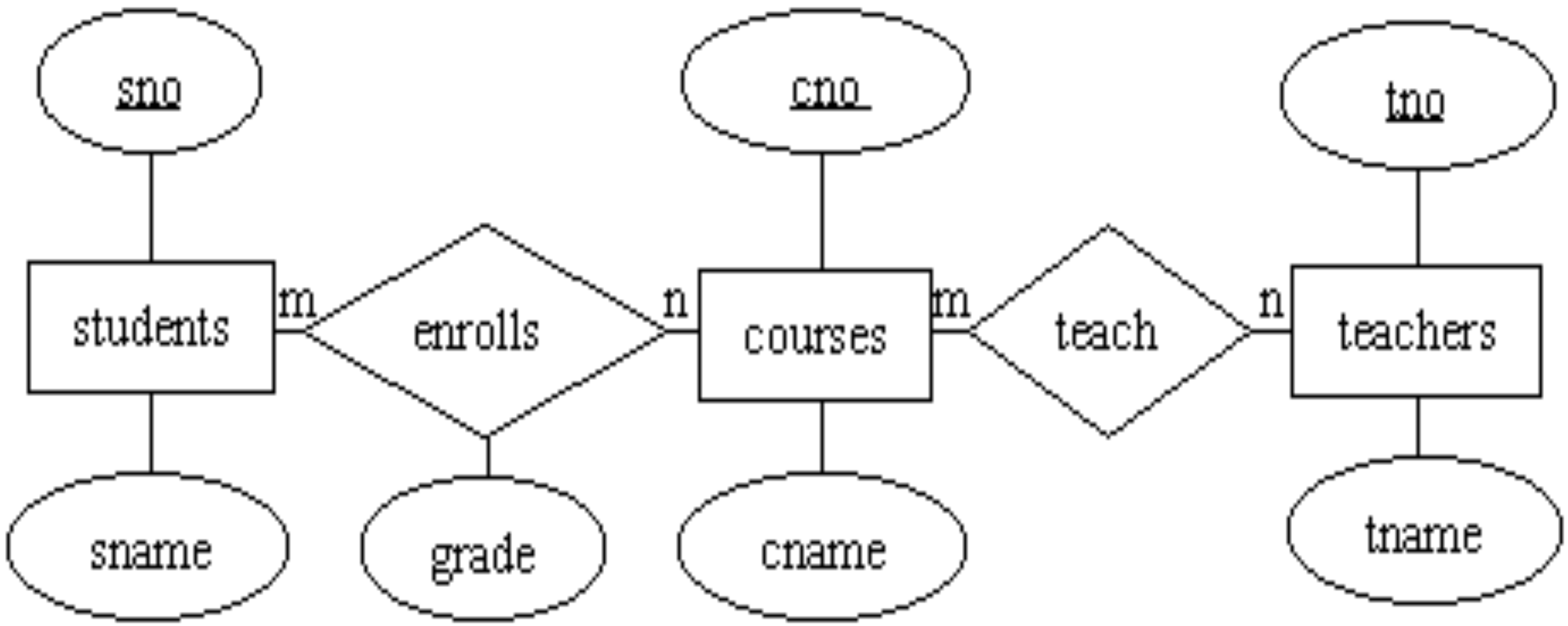
关系数据库设计的核心：

关系模式设计

按照一定的原则从数量众多而又相互关联的数据中，构造出一组既能较好地**反映现实世界**，而又有良好的**操作性能**的关系模式。

关系模式设计问题的提出

例：设计教学管理关系数据库模型



简单教学管理的实体联系模型 ER 图

例：设计教学管理关系数据库模型

解：sct（sno, cno, tno, sname, grade, cname, tname）

关系SCT

Sno	Cno	Tno	Sname	Grade	Cname	Tname
S1	C1	T1	赵民	90	OS	彭
S1	C2	T2	赵民	90	DS	杨
S1	C3	T3	赵民	85	C++	刘
S1	C4	T4	赵民	87	DB	张
S2	C1	T4	李军	90	OS	张
S3	C1	T4	陈江	75	OS	张
S3	C2	T2	陈江	70	DS	杨
S3	C4	T4	陈江	56	DB	张
S4	C1	T1	魏致	90	OS	彭
S4	C2	T2	魏致	85	DS	杨
S5	C1	T1	乔远	95	OS	彭
S5	C4	T4	乔远	80	DB	张

问题分析

- 冗余度高
- 修改困难
- 插入问题
- 删除问题

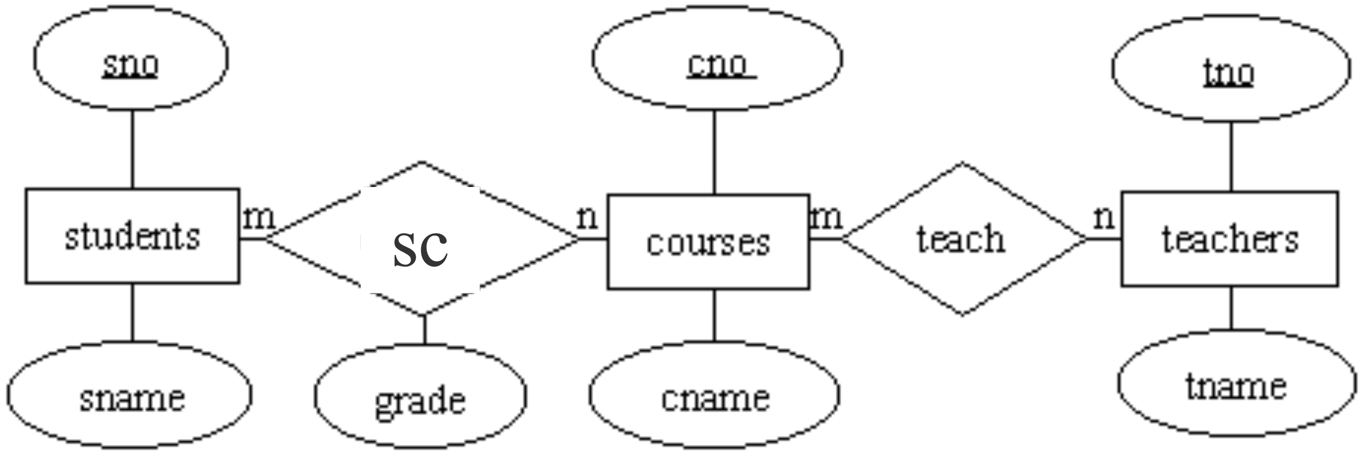
产生问题的原因？

属性间约束关系（即数据间的依赖关系）太强

例：设计教学管理关系数据库模型

解决问题的方法：

分解关系



解： sct（sno, cno, tno, sname, grade, cname, tname)

解一：

students（sno, sname)
courses （cno, tno, cname)
teachers（tno, cno, tname)
sc （sno, cno, grade)

解二：

students（sno, sname)
courses （cno, cname)
teachers（tno, tname)
sc （sno, cno, grade)
teaching（tno, cno)

例：设计教学管理关系数据库模型

Students

Sno	Sname
S1	赵民
S2	李军
S3	陈江
S4	魏致
S5	乔远

Teachers

Tno	Tname
T1	彭
T2	杨
T3	刘
T4	张

sc

Sno	Cno	Grade
S1	C1	90
S1	C2	90
S1	C3	85
S1	C4	87
S2	C1	90
S3	C1	75
S3	C2	70
S3	C4	56
S4	C1	90
S4	C2	85
S5	C1	95
S5	C4	80

Courses

Cno	Cname
C1	OS
C2	DS
C3	C++
C4	DB

Teach

Cno	Tno
C1	T1
C1	T4
C2	T2
C3	T3
C4	T4

本章要解决的主要问题

- 1. 怎样评价关系模式的优劣
- 2. 怎样将关系模式分解为一组较理想的关系模式

本章内容

- 第一节　问题的提出
- 第二节　函数依赖
- 第三节　范式
- 第四节　数据依赖的公理系统

第二节　函数依赖

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，分函数依赖和多值依赖。

函数依赖(Functional Dependency, FD)是数据依赖的一种，它反映属性或属性组之间相互依存，互相制约的关系，即反映现实世界的约束关系。

（1） 函数依赖定义

设R（U）是属性集U上的一个关系模式, X和Y均为$U=\{A_1, A_2, \ldots, A_n\}$的子集, r为R的**任一关系**, 如果对于r中的**任意两个**元组u, v, 只要有u[X] = v[X], 就有u[Y] = v[Y], 则称**X函数决定Y**, 或称**Y函数依赖于X**, 记为**X→Y**。

“任一、任意两个”： 表明具体的函数依赖是基于该模式的所有可能数据推导出来，或所有数据必须符合这个函数依赖

函数依赖

例： 1. {sno}→{sname},

{cno}→{cname},

{sno, cno}→{grade} ✓

2. {tno}→{cno},

{sno}→{tname} ✗

只能根据语义来确定函数依赖性的存在与否。

(2) 属性间的联系决定函数依赖关系

设X、Y均是U的子集

1. X和Y间联系是1:1，则X→Y，Y→X。

2. X和Y间联系是M:1(M>1)，则X→Y。

3. X和Y间联系是M:N(M，N>1)，则X、Y间不存在函数依赖。

函数依赖举例

SCX(SNO, SNAME, AGE, SEX, CNO, CNAME, GRADE)

X Y

SNO → SNAME

SNO → AGE

SNO → SEX

(SNO, CNO) → GRADE

CNO → CNAME

SNO $\not\rightarrow$ GRADE

(3) 自反性、增广性和传递性

设U是关系模式R的属性集，F是R上成立的只涉及到U中属性的函数依赖集，则：

自反性：若$Y \subseteq X \subseteq U$，则$X \rightarrow Y$在R上成立

增广性：若$X \rightarrow Y$在R上成立，且$Z \subseteq U$，则$XZ \rightarrow YZ$在R上成立

传递性：若$X \rightarrow Y$和$Y \rightarrow Z$在R上成立，则$X \rightarrow Z$在R上成立

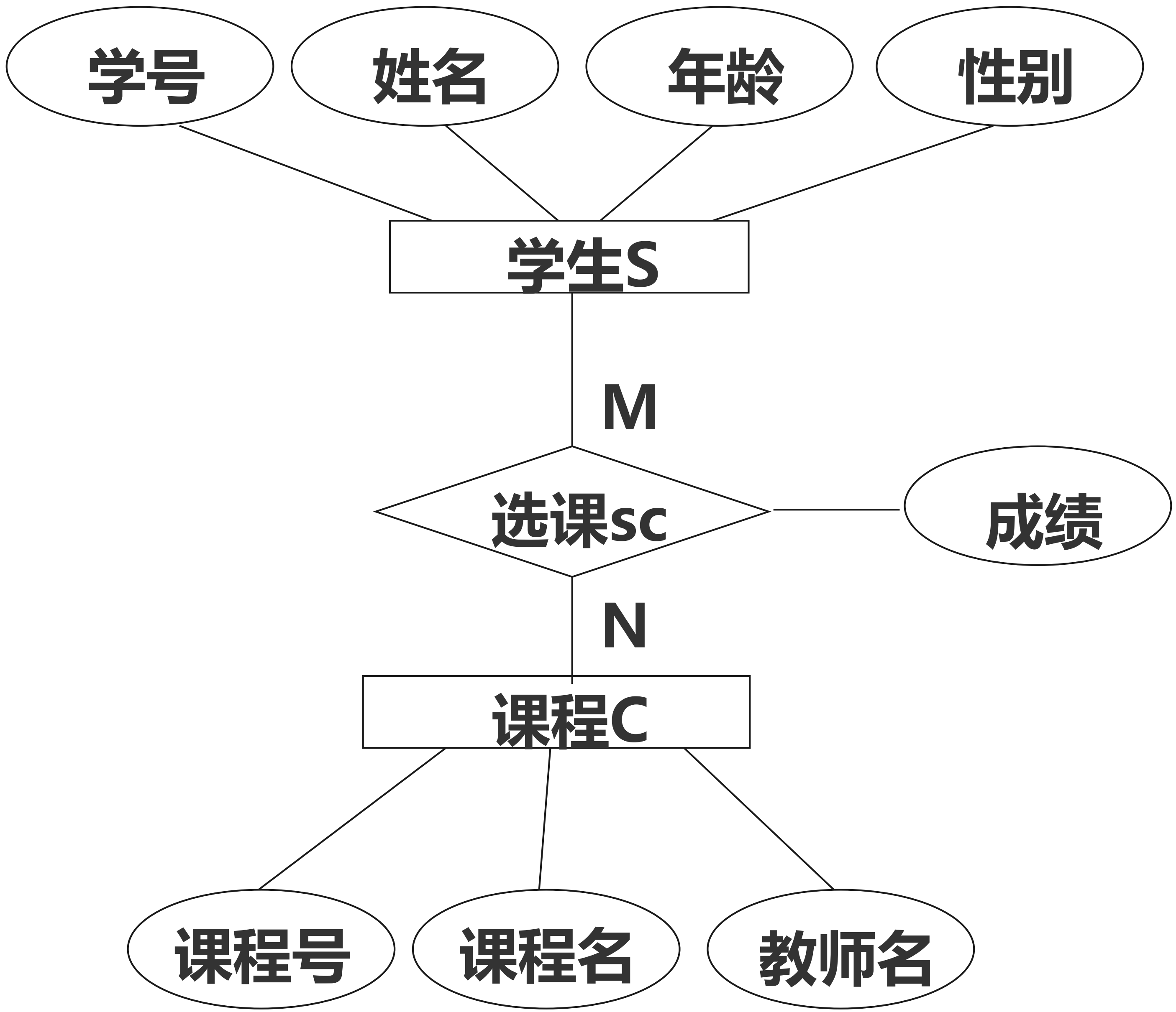
(4)完全函数依赖和部分函数依赖

在R(U)中，如果X→Y，并且对于X的任何真子集X'都不存在X'→Y，则称**Y完全函数依赖于X**，记作$X \xrightarrow{F} Y$；否则，如果X→Y，且X中存在一个真子集X'，使得X'→Y成立，则**Y部分函数依赖于X**，记作$X \xrightarrow{P} Y$。

函数依赖

例如：

学号—SNO
姓名—SN
年龄—age
性别—sex
课程号—CNO
课程名—CN
教师名—TN
成绩—G



完全与部分函数依赖举例

例： $(sno, cno) \xrightarrow{P} cn$

$(sno, cno) \xrightarrow{F} tn$

$(cno) \xrightarrow{F} cn$

$(sno, cno) \xrightarrow{F} g$

(5) 平凡函数依赖和非平凡函数依赖

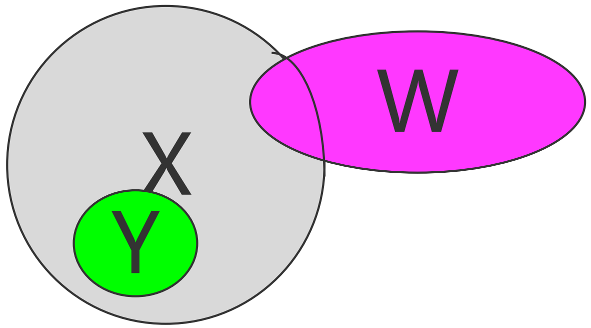
设X, Y均为某关系上的属性集，且X→Y

1）若Y包含于X，则称X→Y为：**平凡函数依赖**；

2）若Y不包含于X，则称X→Y为：**非平凡函数依赖**。
(一般都指的是非平凡函数依赖)

平凡与非平凡函数依赖举例

例： 设X，Y，W为关系R中的三个属性组，属性关系如下图所示，问X→Y，X→W，W→Y各属于上述何种函数依赖。



解：X→Y　　　　为平凡函数依赖

X→W，W→Y　　为非平凡函数依赖

本章内容

第一节　问题的提出

第二节　函数依赖

第三节　范式

第四节　数据依赖的公理系统

第三节　范式(normal form)

关系模式满足的确定约束条件称为**范式**，根据满足约束条件的级别不同，范式由低到高分为：1NF, 2NF, 3NF, BCNF, 4NF, 5NF等。

关系模式的规范化：

把一个低一级范式的关系模式**分解**为高一级范式的关系模式的过程。

1.第一范式（1NF)

关系模式的所有域为简单域，其元素不可再分，即属性不能再分。

例1： $A_1, A_2, A_3, \ldots, A_k, \ldots, A_n$

$$\begin{array}{c} A_k \\ \swarrow \quad \searrow \\ A_{k1} \quad A_{k2} \end{array}$$

例2： 工资(工号, 姓名, 工资(基本工资, 年绩津贴, 煤电补贴))

1.第一范式（1NF)

★不满足1NF的关系称为**非规范化关系**。

★关系数据模型不能存储上述两个例子（非规范化关系）

★转化方法：

1) $A_1, A_2, A_3, \ldots, A_{k1}, A_{k2}, \ldots, A_n$

2)工资(工号, 姓名, 基本工资, 年绩津贴, 煤电补贴)

1.第一范式（1NF)

- 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

第一范式（1NF）举例

例:关系模式 S-L-C(Sno, Sdept, Sloc, Cno, Grade)

Sloc为学生住处，假设每个系的学生住在同一个地方

存在的函数依赖包括：

Sno → Sdept

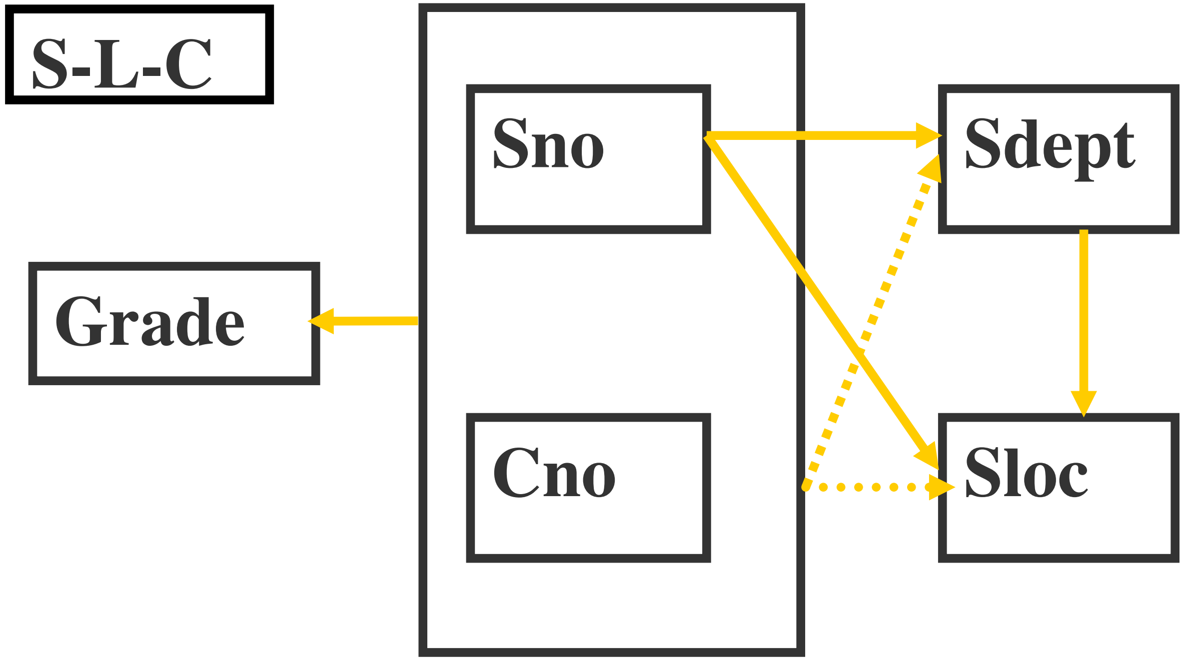
Sno → Sloc

Sdept → Sloc

$(Sno, Cno) \xrightarrow{F} Grade$

$(Sno, Cno) \xrightarrow{P} Sdept$

$(Sno, Cno) \xrightarrow{P} Sloc$



虚线表示部分函数依赖

S-L-C的码为(Sno, Cno)、S-L-C满足第一范式。

非主属性Sdept和Sloc部分函数依赖于码(Sno, Cno)

存在的问题

S-L-C不是一个好的关系模式

S-L-C(Sno, Sdept, Sloc, Cno, Grade)

(1) 插入异常

(2) 删除异常

(3) 数据冗余度大

(4) 修改复杂

原因：存在部分函数依赖

S-L-C不是一个好的关系模式

- 原因

  Sdept、 Sloc部分函数依赖于码。

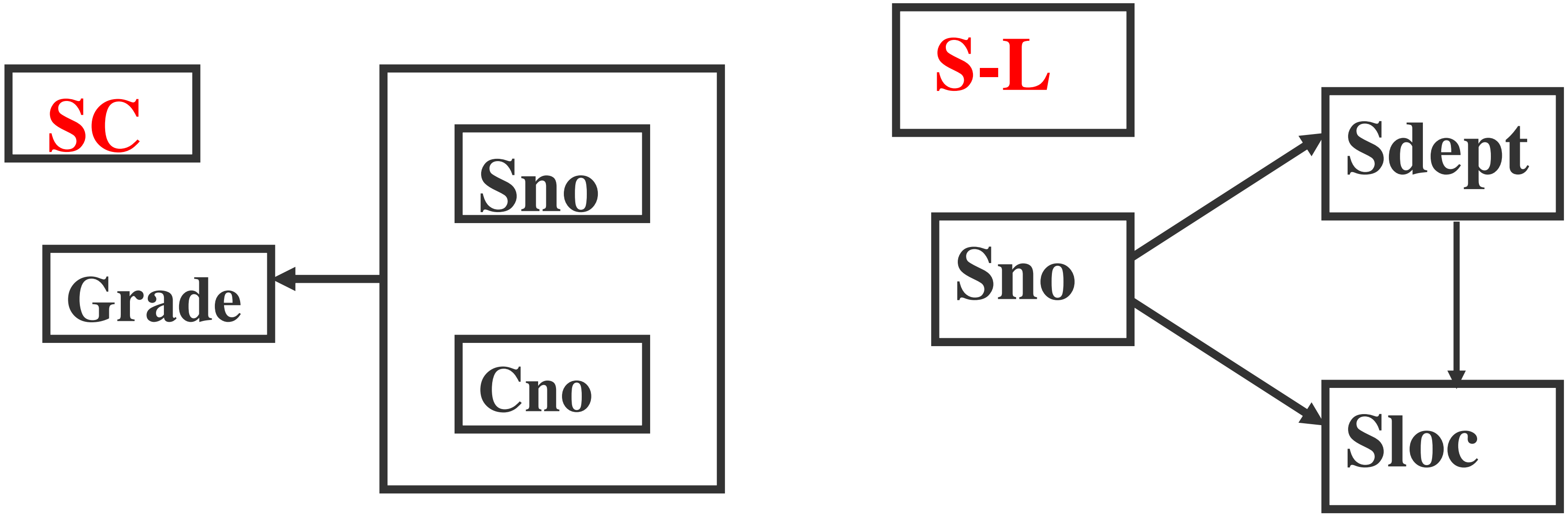
- 解决方法

  S-L-C (Sno, Sdept, Sloc, Cno, Grade)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）

消除部分函数依赖



- 关系模式SC的码为（Sno, Cno）
- 关系模式S-L的码为Sno
- 这样**非主属性对码都是完全函数依赖**

2.第二范式（2NF）

给定关系模式R及其上的函数依赖集F，如果R的任何一个非主属性都完全函数依赖于码，则称R满足第二范式，简记为R∈2NF。

2.第二范式（2NF)

例1：

S-L-C(Sno, Sdept, Sloc, Cno, Grade) $\in$ 1NF

S-L-C(Sno, Sdept, Sloc, Cno, Grade) $\notin$ 2NF

SC(Sno, Cno, Grade) $\in$ 2NF

S-L(Sno, Sdept, Sloc) $\in$ 2NF

举例：判定是否2NF

例2：

SCT（SNO, CNO, CN, GRADE, TNAME, SALARY）

F={(SNO, CNO)→GRADE, CNO→CN,
CNO→TNAME, TNAME→SALARY}

//一门课只能由一个老师教；教师名唯一

候选码:　　(sno, cno)

是否存在非主属性对码的不完全依赖?

存在：$(SNO, CNO) \xrightarrow{P} CN$

$(SNO, CNO) \xrightarrow{P} TNAME$

$(SNO, CNO) \xrightarrow{P} SALARY$

答案：非2NF

2.第二范式（2NF)

解决的方法：是用**投影分解**把关系模式SCT分解为两个关系模式

SCT（<u>SNO, CNO</u>, CN, GRADE, TNAME, SALARY）

分解：SC(<u>SNO, CNO</u>, GRADE),

CT(<u>CNO</u>, CN, TNAME, SALARY)

分解后得到的两个关系模式均为2NF

2NF存在的问题

分解为2NF后，是否都不存在前述的四种问题？

回顾例1 S-L-C(Sno, Sdept, Sloc, Cno, Grade) 分解为：

SC（<u>Sno, Cno</u>, Grade） ∈ 2NF

S-L（<u>Sno</u>, Sdept, Sloc） ∈ 2NF

在SC中，不存在前述的四种问题；

但在S-L中，仍然存在着前述的四种问题.

说明2NF还不够好。

2NF存在的问题

回顾例2

SCT（<u>SNO, CNO</u>, CN, GRADE, TNAME, SALARY）

分解为：

SC(<u>SNO, CNO</u>, GRADE),

CT(<u>CNO</u>, CN, TNAME, SALARY)

在SC中，不存在前述的四种问题；

但在CT中，仍然存在着前述的四种问题.

说明2NF还不够好。

范式中存在的传递函数依赖

S-L(<u>Sno</u>, Sdept, Sloc)
CT(<u>CNO</u>, CN, TNAME, SALARY)

Sno → Sdept, Sdept → Sloc

CNO→TNAME, TNAME→SALARY

传递函数依赖

在R(U)中，如果X→Y, Y→Z, 并且$Y \not\rightarrow X$, 那么称X→Z是**传递函数依赖**，即**Z传递函数依赖于X**。

3. 第三范式（3NF)

给定关系模式R及其上的函数依赖集F, 如果R的任何一个非主属性都不传递函数依赖于它的任何一个侯选码, 则称R满足第三范式, 简记为R∈3NF。

定理：一个3NF的关系(模式)必定是2NF的(3NF⊆2NF⊆1NF) 。

第三范式（3NF）举例

例1： CT（<u>CNO</u>, CNAME, TNAME, SALARY）

F={ CNO→TNAME, CNO→CNAME,
TNAME→SALARY}

问题：

插入和删除异常

原因：

存在非主属性SALARY传递函数依赖于码，故 CT∈2NF，不是3NF.

分解：

C(CNO, CNAME, TNAME),
T(TNAME, SALARY)均是3NF

3. 第三范式（3NF）

例 2： SP={ <u>SNO</u>, SNAME, CITY, province }
F={ SNO→SNAME, SNO→CITY,
CITY→province}

存在非主属性province对候选码的传递函数依赖, 故SP为2NF, 不是3NF。

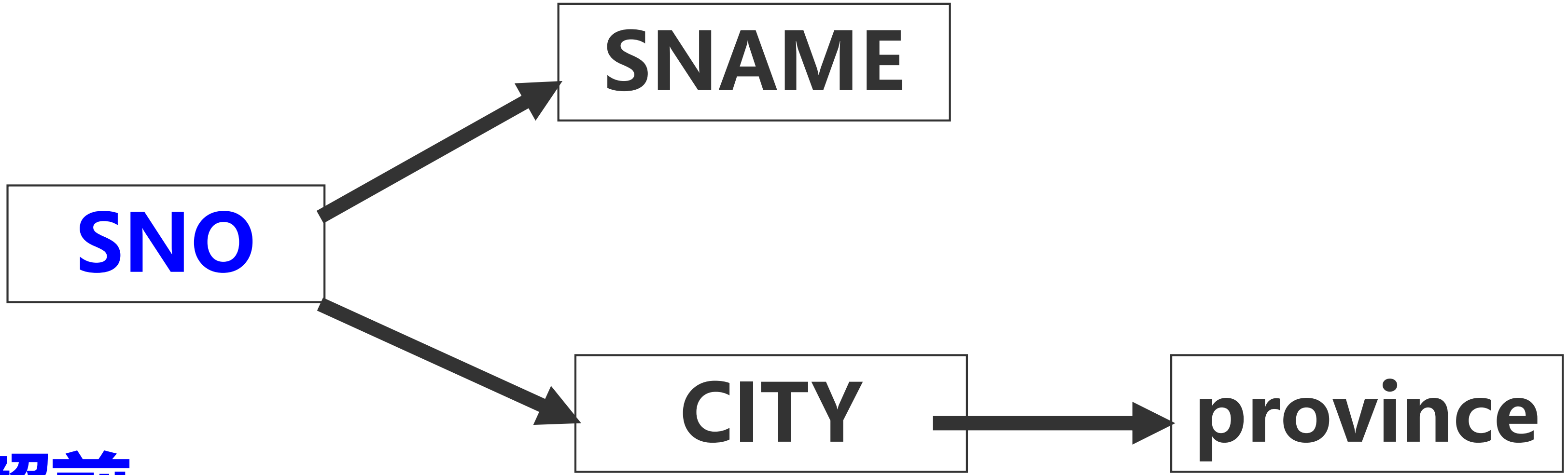
分解:

SP1={ <u>SNO</u>, SNAME, CITY }

SP2={ <u>CITY</u>, province }

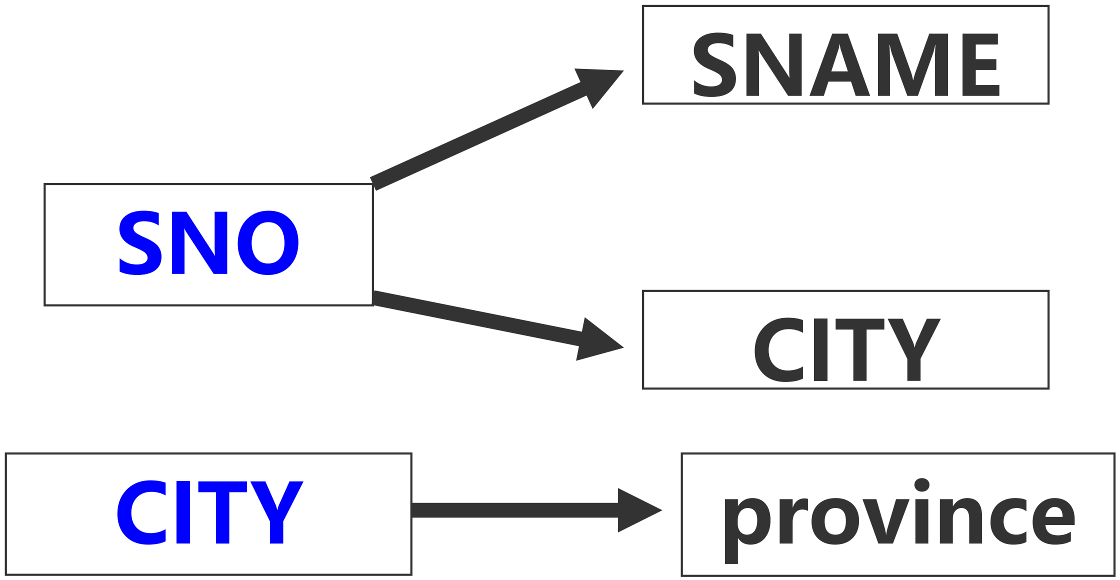
分解后均为3NF。

第三范式（3NF）分解举例



分解前

分解后

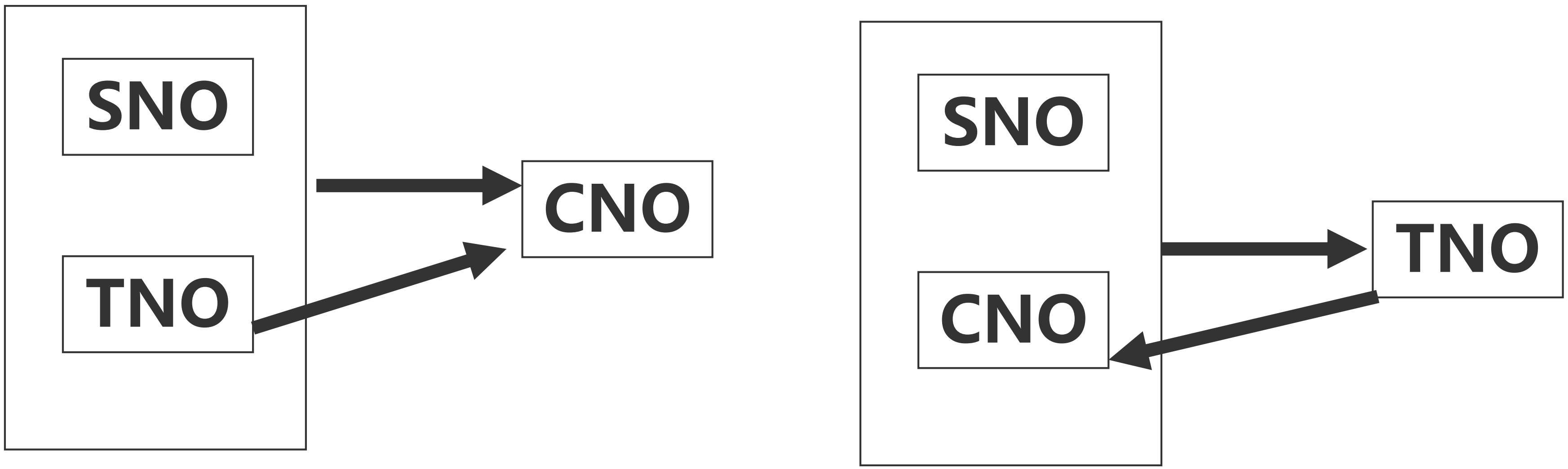


第三范式（3NF）举例

例3： R={SNO, TNO, CNO}

F={TNO→CNO, (SNO, TNO)→CNO, (SNO, CNO)→TNO}

//一名教师只能上一门课；一门课可由多个教师上



候选码为（SNO, TNO）和（SNO, CNO）

R中没有非主属性, 故R为3NF。

第三范式（3NF）不一定就没问题

3NF是否就一定不存在以前的四种问题呢?

回顾例 3：

R={SNO, TNO, CNO}

F={TNO→CNO, (SNO, CNO)→TNO, (SNO, TNO)→ CNO}

候选码为（SNO, TNO）和（SNO, CNO）

仍然存在问题

例如要插入某个教师能讲授某门课的信息, 而又没有学生选修该课时, 会有插入异常。

4.BCNF（改进的3NF)

给定关系模式R及其上的函数依赖集F，如果F中**每个非平凡函数依赖**X→Y的左部（决定因素）X 中 必 含 有 候 选 键 （ 码 ）， 则 称 R 满 足 Boyce/Codd 范式，简记为R∈ BCNF

R∈1NF，若X→Y且Y ⊈ X，X中必含有候选键（码），则R∈BCNF

4.BCNF（改进的3NF）

BCNF的内涵：

1.非主属性对码完全函数依赖

2.非主属性不传递函数依赖于任何一个码；

3.主属性对不含它的码完全函数依赖；

4.主属性不传递函数依赖于任何一个码；

5.没有属性完全函数依赖于一组非主属性；

定理： BCNF满足3NF

(BCNF $\subseteq$ 3NF $\subseteq$ 2NF $\subseteq$ 1NF")

4.BCNF（改进的3NF）

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

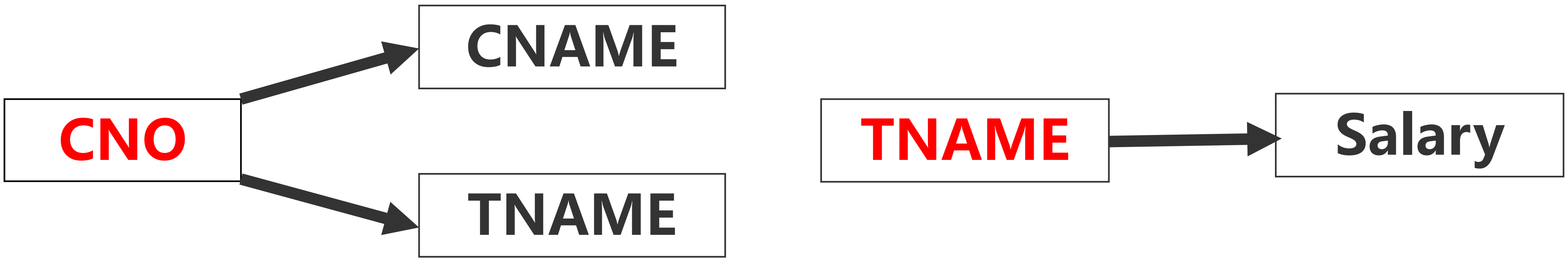
↓ 消除主属性对码的部分和传递函数依赖

BCNF

BCNF（改进的3NF）举例

例1： C(<u>CNO</u>, CNAME, TNAME),

T(<u>TNAME</u>, SALARY)

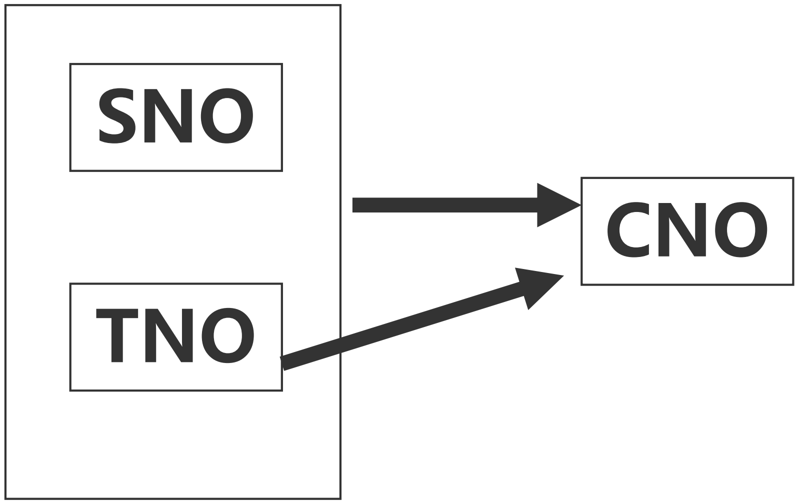


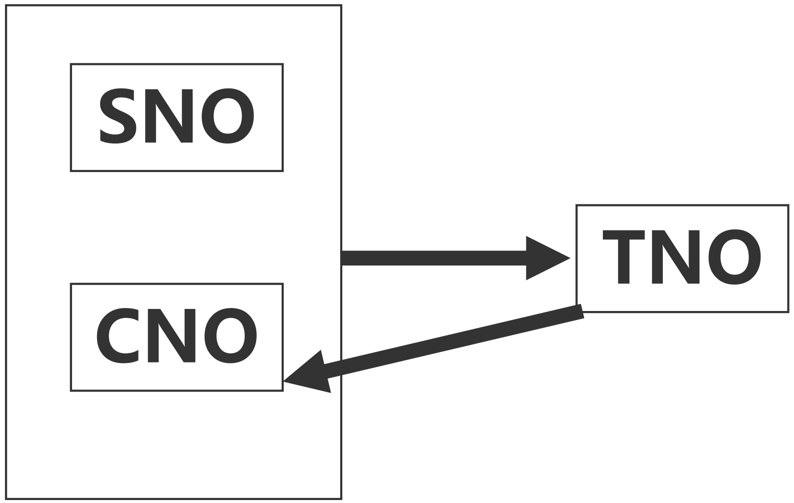
可以由每个函数依赖的左部均包含码，
判断出C和T均属于BCNF。

BCNF（改进的3NF）举例

例2： R={SNO, TNO, CNO}

F={TNO→CNO, (SNO, TNO)→CNO, (SNO, CNO)→TNO}





由TNO→CNO知, **R不是BCNF**。

分解： ST(SNO, CNO), TC(TNO, CNO)

5.多值依赖

函数依赖有效地表达了属性之间的多对一联系，但不能表达属性之间的一对多联系，更不能表达属性之间的多对多联系。

5.多值依赖

例1：

课程C	教师T	参考书B
物理	{李勇，王军}	{普通物理学，光学原理，物理习题集}
数学	{李勇，张平}	{数学分析，微分方程，高等代数}
计算数学	{张平，周峰}	…

课程C	教师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

该模式的候选码是 (C, T, B)，即全码，

是BCNF，但依然存在各种异常。

1)多值依赖的定义

设R（U）是属性U上的一个关系模式，X，Y，Z均为U={A_1，A_2，…，A_n}的子集，并且Z=U-X-Y，用小写字母x，y，z分别代表属性集X，Y，Z 的值，关系模式R（U）中多值依赖X→→Y成立，当且仅当对R(U)的任一关系r，给定一对(x，z)值，有一组Y的值，这组值仅仅决定于x值而与z值无关。

多值依赖案例

例1：

课程C	教师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

例1中存在如下多值依赖：　C→→T, C→→B

多值依赖案例

例2： 存储（仓库W, 保管员S, 货物C）

仓库和保管员　1：n

仓库和货物　　m:n

保管员和货物　m:n

其候选码为（W, S, C）

W→→S, W→→C

W	S	C
w1	s1	c1
w1	s1	c2
w1	s2	c1
w1	s2	c2
w2	s3	c1
w2	s3	c3
w2	s4	c1
w2	s4	c3

➢ 多值依赖的性质

（1）多值依赖具有对称性。即若X→→Y，则X→→Z，其中Z=U-X-Y。

（2）多值依赖的传递性。即若X→→Y，Y→→Z，则X→→Z-Y

（3）函数依赖可以看作是多值依赖的特殊情况，即若X→Y，则X→→Y。

（4）若X→→Y，X→→Z，则X→→YZ

（5）若X→→Y，X→→Z，则X→→Y∩Z

（6）若X→→Y，X→→Z，则X→→Y-Z，X→→Z-Y

2)函数依赖和多值依赖的区别

(1)函数依赖在小范围内成立, 则在大范围上一定成立；多值依赖在小范围内成立, 在更大范围上未必成立。

例如：S(SNO, SN, SEX, AGE)

函数依赖SNO→SN在小范围(SNO, SN, SEX)上成立, 它在大范围(SNO, SN, SEX, AGE)上也成立;

多值依赖是否成立与属性函数有关

又如：

(仓库W, 保管员S, 货物C, 保管员的地址L)

在(W, S, C)上有多值依赖W→→S, W→→C成立

多值依赖W→→S在小范围(W, S, C)内成立, **但在大范围(W, S, C, L)上就不成立。**

多值依赖W→→C在小范围(W, S, C)内成立, 在大范围(W, S, C, L)上也成立。

W	S	C	L
w1	s1	c1	a1
w1	s1	c2	a1
w1	s2	c1	a2
w1	s2	c2	a2
w2	s3	c1	a3
w2	s3	c3	a3
w2	s4	c1	a4
w2	s4	c3	a4

5.多值依赖

(2) 若函数依赖X→Y在R(U)上成立, 则对于任何Y的真子集Y’均有X→Y’也成立；对于多值依赖X→→Y, 若在R(U)上成立, 却不能断言对于任何Y的真子集Y’均有X→→Y’也成立。

3)平凡的多值依赖和非平凡的多值依赖

对于属性集U上的一个多值依赖
X→→Y （X, Y均为U的子集），如果X
包含Y或者 Z=U-X-Y= Φ, 则称
X→→Y为**平凡的多值依赖**, 否则称
X→→Y为**非平凡的多值依赖**。

6.第四范式

设R是一个关系模式，D是R上的多值依赖集合。如果D中成立非平凡多值依赖X→→Y时，X必包含R的码，那么称R是**第四范式的模式**，简记为R∈4NF。

如果关系模式是**4NF**，则必为**BCNF**

例如：存储（仓库W，保管员S，货物C）

候选码是**(W, S, C)**

多值依赖W→→S，W→→C都是非平凡的，且决定因素都不包含候选码，**故它不是4NF**

6.第四范式

例如：

课程C	教师T	参考书B
物理	{李勇，王军}	{普通物理学，光学原理，物理习题集}
数学	{李勇，张平}	{数学分析，微分方程，高等代数}
计算数学	{张平，周峰}	…

课程C	教师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

候选码是（C, T, B)，存在多值依赖C→→T，C→→B

不是4NF

6.第四范式

一个关系模式达到BCNF，但不是4NF，仍然有不好的性质：

数据冗余度太大，插入删除复杂。

采用**投影分解**的方法消去非平凡且非函数依赖的多值依赖，将CTB分解为**CT**（C, T），**CB**（C, B），这都是平凡的多值依赖，所以**CT、CB是4NF**。

规范化小结

- 关系数据库的规范化理论是数据库逻辑设计的工具
- 目的：尽量消除插入、删除异常, 修改复杂, 数据冗余
- 基本思想：逐步消除数据依赖中不合适的部分
  - 实质：概念的单一化

关系模式规范化的基本步骤

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

规范化小结（续）

- 不能说规范化程度越高的关系模式就越好

- 在设计数据库模式结构时，必须对现实世界的实际情况和用户应用需求作进一步分析，确定一个合适的、能够反映现实世界的模式

- 上面的规范化步骤可以在其中任何一步终止

例1：

设有一个关系模式R(A, B, C, D), F是R上成立的函数依赖, F={A→B, A→C}, 试写出关系模式R的候选键, 分析其满足的范式并将其分解为满足BCNF的关系模式。

解： R的候选码为(A, D)

R为1NF, 因为存在着非主属性B, C对码的部分函数依赖（不是2NF）。

可分解为:（A, D）和（A, B, C）均为BCNF

例2：

已知关系模式学生(学号, 姓名, 系名, 系主任, 课程, 成绩), 其上成立的函数依赖有：
{学号→姓名, 学号→系名, 系名→系主任, （学号, 课程）→成绩}
试确定该关系模式的候选键及满足第几范式。

其候选码为 (学号, 课程)

该关系模式为1NF, 因为存在着非主属性对码的部分函数依赖, 所以不是2NF 。

例2

- **将其分解为2NF**

R1(<u>学号</u>, 姓名, 系名, 系主任)

学号→姓名, 学号→系名, 系名→系主任,

R2(<u>学号, 课程</u>, 成绩) （学号, 课程）→成绩

- **将其分解为BCNF**

R1(<u>学号</u>, 姓名, 系名)

学号→姓名, 学号→系名

R2(<u>系名</u>, 系主任)

系名→系主任

R3(<u>学号, 课程</u>, 成绩)

（学号, 课程）→成绩

本章内容

第一节　问题的提出

第二节　函数依赖

第三节　范式

第四节　数据依赖的公理系统

第四节 数据依赖的公理系统

一、函数依赖的逻辑蕴涵

定义1：F逻辑蕴含X→Y

对于关系模式R〈U, F〉，其任何一个具体关系r, 若函数依赖X→Y都成立, 则称<u>函数依赖集F逻辑蕴含X→Y</u>

tips：即能从F中推导出X->Y

定义2：函数依赖集F的闭包

在关系模式R〈U, F〉中为F所逻辑蕴含的函数依赖的全体叫做<u>F的闭包</u>, 记作F+

tips：即能从函数依赖集F中推导出的所有函数依赖组成的集合

一、函数依赖的逻辑蕴涵

例1：

关系模式 R=(A, B, C), 函数依赖集F={A→B, B→C},
F逻辑蕴涵A→B, B→C, A→C。

一、函数依赖的逻辑蕴涵

例2：R=（A, B, C), F={A→B, B→C}, 求F^+

F^+ ={A→Φ, AB→Φ, AC→Φ, ABC→Φ, B→Φ, C→Φ

A→A, AB→A, AC→A, ABC→A, B→B, C→C,

A→B, AB→B, AC→B, ABC→B, B→C,

A→C, AB→C, AC→C, ABC→C, B→BC,

A→AB, AB→AB, AC→AB, ABC→AB, BC→Φ,

A→AC, AB→AC, AC→AC, ABC→AC, BC→B,

A→BC, AB→BC, AC→BC, ABC→BC, BC→C, Φ→Φ

A→ABC, AB→ABC, AC→ABC, ABC→ABC, BC→BC }

二、Armstrong公理

定理：若U为关系模式R的属性全集，F为U上的一组函数依赖，设X、Y、Z、W 均为U的子集，对R(U, F)有：

F1(自反律)：若Y $\subseteq$ X(表示X包含Y），则 X→Y为F所蕴涵；

F2(增广律)：若X→Y为F所蕴涵，则 XZ→YZ为F所蕴涵；

F3(传递律)：若X→Y，Y→Z为F所蕴涵，则 X→Z为F所蕴涵；

定理 6.1 Armstrong推理规则是正确的

(l) **自反律**: 若Y ⊆ X ⊆ U, 则X →Y 为F 所蕴含

证: 设Y ⊆ X ⊆ U

对R <U, F> 的任一关系r 中的任意两个元组t, s:

若t [X]=s[X], 由于Y ⊆ X, 有t [y]=s [y],

所以X →Y 成立, 自反律得证

(2)**增广律**: 若X→Y为F所蕴含, 且Z ⊆ U, 则XZ→YZ 为F所蕴含。

证：设X→Y为F所蕴含, 且Z ⊆ U。

设R<U, F> 的任一关系r中任意的两个元组t, s:

若t [XZ]=s[XZ], 则有t [X]=s[X]和t [Z]=s[Z];

由X→Y, 于是有t [Y]=s[Y], 所以t [YZ]=s[YZ], 所以XZ→YZ为F所蕴含, 增广律得证。

定理 6.l Armstrong推理规则是正确的（续）

(3) **传递律**：若X→Y及Y→Z为F所蕴含，则 X→Z为 F所蕴含。

证：设X→Y及Y→Z为F所蕴含。

对R<U, F> 的任一关系 r中的任意两个元组 t, s：

若t [X]=s[X]，由于X→Y，有 t [Y]=s[Y]；

再由Y→Z，有t [Z]=s[Z]，所以X→Z为F所蕴含，传递律得证。

导出规则

1.根据A1, A2, A3这三条推理规则可以得到下面三条推理规则：

- 合并规则：由$X \to Y$, $X \to Z$, 有$X \to YZ$。
  (A2, A3)　　tips：根据A2可得， X->XY, XY->YZ.再根据A3，X->YZ
- 伪传递规则：由$X \to Y$, $WY \to Z$, 有$XW \to Z$。
  (A2, A3)
- 分解规则：由$X \to Y$及 $Z \subseteq Y$, 有$X \to Z$。
  (A1, A3)

2.根据合并规则和分解规则, 可得引理6.1

引理6.l　$X \to A_1 A_2 \ldots A_k$成立的充分必要条件是$X \to A_i$成立（$i=l, 2, \ldots, k$）

Armstrong公理系统

- Armstrong公理系统是有效的、完备的
  - 有效性：

    由F出发根据Armstrong公理推导出来的每一个函数依赖一定在F+中；
  - 完备性：

    F+中的每一个函数依赖，必定可以由F出发根据Armstrong公理推导出来
- 有效性的证明可通过Armstrong推理规则的正确性得到；要证明完备性，首先要解决如何判定一个函数依赖是否属于由F根据Armstrong公理推导出来的函数依赖的集合

三、属性集闭包

定义：若F为关系模式R(U)的函数依赖集, X是U的子集, 则由Armstrong公理推导出的所有X→Ai所形成的属性集{Ai|i=1, 2, …}称为**X关于F的闭包**, 记为**XF+**。

tips：与闭包F+的区别。闭包集合的元素是函数依赖，属性集闭包的元素是属性

属性集闭包

例1：设R=(A, B, C), F={A→B, B→C}
当X分别为A, B, C时, 求XF+ 。

解：1)当X=A时,
XF+ = {A, B, C}
2)当X=B时,
XF+ = {B, C}
3)当X=C时,
XF+ = {C}

关于闭包的引理

- **引理6.2**

  设F为属性集U上的一组函数依赖, X, Y ⊆ U, X→Y能由F 根据Armstrong公理导出的充分必要条件是Y ⊆XF+

  tips: 只有Y这个属性在X关于F的属性集闭包中，才能推出X->Y。只要能推导出X->Y，那么Y一定在X关于F的属性集闭包中。

- **用途**

  将判定X→Y是否能由F根据Armstrong公理导出的问题, 转化为求出XF+ 、判定Y是否为XF+的子集的问题

求闭包的算法

算法6.1 求属性集X（X ⊆ U）关于U上的函数依赖集F

的闭包XF+

输入：X, F　　　　**输出**：XF+

步骤:

（1）令X（0）=X, i=0

（2）求B, 这里B = { A |(∃ V)(∃ W)
(V→W∈F∧V ⊆ X（i）∧A∈ W)};

即找出F中左边为X(i)子集的函数依赖，其中B为所有函数依赖中未出现在X(i)中的属性

（3）X（i+1）=B∪X（i）

（4）判断X（i+1）= X （i）吗?

（5）若相等或X（i）=U , 则X（i）就是XF+, 算法终止。

（6）若否, 则 i=i+l, 返回第（2）步。

求闭包的算法

对于算法6.1, 令ai =|X（i）|,

{ai }形成一个步长大于1的严格递增的序列,

序列的上界是 | U |,

因此该算法最多 |U| - |X| 次循环就会终止。

函数依赖闭包

[例1] 已知关系模式R<U, F>, 其中

U={A, B, C, D, E};

F={AB→C, B→D, C→E, EC→B, AC→B}。

求（AB）F+ 。

解 设X（0）=AB;

(1)X（1）=AB∪CD=ABCD。

(2) X（0）≠ X（1）

X（2）=X（1）∪BE=ABCDE。

(3) X（2）=U, 算法终止

→ （AB）F+ =ABCDE。

四、函数依赖集的等价与覆盖

定义

设F和G是关系模式R(U)上的两个函数依赖集，如果$F^+ = G^+$，则称**F等价于G**，亦称F覆盖G，或G覆盖F，记为F≡G。

五、最小函数依赖集

设F为函数依赖集，如果F满足：

(1)F中每个函数依赖X→Y的右边Y均为单个属性；

(2)F中任何一个函数依赖X→A，其F-{X→A}与F都不等价；

tips：F中无冗余的函数依赖

(3)F中任何一个函数依赖X→A，Z为X的真子集，(F-{X→A})∪{Z→A}与F都不等价；

tips：F中函数依赖的左端无冗余属性

则称F为极小函数依赖集（最小依赖集或最小覆盖），记作Fmin 。

最小函数依赖集

[例] **关系模式R<U, F>, 其中:**

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

F ={ Sno→Sdept, Sdept→Mname,

(Sno, Cno)→Grade }

设F’ ={Sno→Sdept, Sno→Mname, Sdept→Mname,

(Sno, Cno)→Grade, (Sno, Sdept)→Sdept}

F是最小覆盖, 而F’ 不是。

因为：F ’ - {Sno→Mname}与F ’ 等价

F ’ - {(Sno, Sdept)→Sdept}也与F ’ 等价

六、 极小化过程

定理6.3 每一个函数依赖集F均等价于一个极小函数依赖集Fm。此Fm称为F的最小依赖集。

证明: 构造性证明, 找出F的一个最小依赖集。

(1)逐一检查F中各函数依赖FDi：X→Y,
若Y=A1A2 ...Ak, k > =2,
则用 { X→Ai|i=1, 2, ..., k} 来取代X→Y。

tips:将F中的所有函数依赖的右边化为单一属性

(2)逐一检查F中各函数依赖FDi：X→A, 令G=F-{X→A},
若A∈XG+, 则从F中去掉此函数依赖。

tips:去掉F中所有冗余的函数依赖

(3)逐一取出F中各函数依赖FDi：X→A, X=B1B2...Bm,
逐一考查Bi （i=l, 2, ..., m）, 若A ∈（X-Bi ）F+ ,
则以X-Bi 取代X。

tips:去掉F中的所有函数依赖左边的冗余属性

极小化过程

[例1]　F = {A→B, B→A, B→C, A→C, C→A}

Fm1、Fm2都是F的最小依赖集：

Fm1= {A→B, B→C, C→A}

Fm2= {A→B, B→A, A→C, C→A}

- F的最小依赖集Fm不唯一
- 极小化过程(定理6.3的证明)也是检验F是否为极小依赖集的一个算法

极小化过程

[例2] 设有一个记录各个球队球员每场比赛进球的关系模式R(球员编号, 比赛场次, 进球数, 球队名, 队长名), 如果规定, 每个队员只能属于一个球队, 每个球队只有一个队长。

(1)试写出该关系模式的基本函数依赖和候选键。

(2)说明R不是2NF模式的理由, 并把R分解为2NF模式集。

(3)进而把R分解为3NF模式集, 并说明理由。

极小化过程［例2］

R(球员编号, 比赛场次, 进球数, 球队名, 队长名), 如果规定, 每个队员只能属于一个球队, 每个球队只有一个队长。

解：其上成立的函数依赖有：

球员编号→球队名, 球队名→队长名,

(球员编号, 比赛场次)→进球数,

其候选码为 (**球员编号, 比赛场次**)

该关系模式为**1NF**

因为存在着非主属性对码的部分函数依赖, 所以不是2NF 。

极小化过程 [例2]

将R(球员编号, 比赛场次, 进球数, 球队名, 队长名)分解为2NF

R1(球员编号, 球队名, 队长名)为2NF

球员编号→球队名, 球队名→队长名

R2(球员编号, 比赛场次, 进球数)为3NF

(球员编号, 比赛场次)→进球数,

将R1分解为3NF

R11(球员编号, 球队名)为3NF

球员编号→球队名

R12(球队名, 队长名)为3NF

球队名→队长名

极小化过程

[例3]：设有关系模式

R(职工号, 项目名, 工资, 部门名, 部门经理),

如果规定, 每个职工可以参加多个项目, 各领一份工资, 每个项目只属于一个部门管理, 每个部门只有一个经理。

(1)试写出该关系模式的基本函数依赖和候选键。

(2)说明R不是2NF模式的理由, 并把R分解为2NF模式集。

(3)进而把R分解为3NF模式集, 并说明理由。

极小化过程 [例3]

R(职工号, 项目名, 工资, 部门名, 部门经理),
每个职工可以参加多个项目, 各领一份工资, 每个项目只属于一个部门管理, 每个部门只有一个经理。

解：其上成立的函数依赖有：

(职工号, 项目名）→工资,

项目名→部门名, 部门名→部门经理

其候选码为 (职工号, 项目名)

该关系模式为1NF

因为存在着非主属性对码的部分函数依赖, 所以不是2NF 。

极小化过程 [例3]

将R(职工号, 项目名, 工资, 部门名, 部门经理)分解为2NF:

R1(项目名, 部门名, 部门经理)为2NF

项目名→部门名, 部门名→部门经理

R2(职工号, 项目名, 工资)为3NF

(职工号, 项目名）→工资

将R1分解为3NF:

R11(项目名, 部门名)为3NF

项目名→部门名

R12(部门名, 部门经理)为3NF

部门名→部门经理