数据库系统概论

An Introduction to Database System

第六章 关系数据理论

©2003-2005，向阳，保留所有权利

第六章 关系数据理论

©2003-2005，向阳，保留所有权利

6.1 问题的提出

关系数据库逻辑设计

- 针对具体问题，如何构造一个适合于它的数据模式
- 数据库逻辑设计的工具——关系数据库的规范化理论

©2003-2005，向阳，保留所有权利

问题的提出

一、概念回顾

二、关系模式的形式化定义

三、关系模式的简化定义

四、什么是数据依赖

五、数据依赖对关系模式影响

© 2003-2005，向阳，保留所有权利

同濟大學 TONGJI UNIVERSITY

一、概念回顾

- 关系
- 关系模式
- 关系数据库

©2003-2005，向阳，保留所有权利

二、关系模式的形式化定义

关系模式由五部分组成，即它是一个五元组：

R(U, D, DOM, F)

R： 关系名

U： 组成该关系的属性名集合

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

©2003-2005，向阳，保留所有权利

关系模式举例

例:

导师和研究生出自同一个域——人，取不同的属性名，并在模式中定义属性向域的映象，即说明它们分别出自哪个域：

DOM（SUPERVISOR-PERSON）

DOM（POSTGRADUATE-PERSON）

- 一个具体的关系模式：

R((导师名、专业名、学生名)，（人、专业名集），（人.导师、人.学生），学生唯一确定导师)

U D DOM F

©2003-2005，向阳，保留所有权利

三、关系模式的简化表示

- 关系模式R（U, D, DOM, F）

  简化为一个三元组：

  R（U, 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

©2003-2005，向阳，保留所有权利

四、什么是数据依赖

1. 数据依赖含义要点

- 一个关系内部属性与属性之间的约束关系
- 现实世界属性值间相互联系的抽象
- 数据内在的性质
- 语义的体现

©2003-2005，向阳，保留所有权利

什么是数据依赖（续）

2. 数据依赖的类型

- 函数依赖（Functional Dependency，简记为FD）

例：

函数依赖极为普遍地存在于现实生活中。比如描述一个学生的关系，可以有学号(Sno)、姓名(Sname)、系名(Sdept)等几个属性。由于一个学号只对应一个学生，一个学生只在一个系学习。因而当“学号”值确定之后，学生的姓名及所在系的值也就被唯一地确定了。属性间的这种依赖关系类似于数学中的函数 $y=f(x)$，自变量 x 确定之后，相应的函数值 y 也就唯一地确定了。

类似的有 Sname = f(Sno)，Sdept = f(Sno)，即 Sno 函数决定 Sname，Sno 函数决定 Sdept，或者说 Sname 和 Sdept 函数依赖于 Sno，记作 Sno→Sname，Sno→Sdept。

©2003-2005，向阳，保留所有权利

什么是数据依赖（续）

- 多值依赖（Multivalued Dependency，简记为MVD

例：

表 6.3 非规范化关系示例

课程 C	教员 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	数学分析 … …
⋮	⋮	⋮

- 其他

©2003-2005，向阳，保留所有权利

五、数据依赖对关系模式的影响

[例1]建立一个描述学校教务的数据库：

学生的学号（Sno）、所在系（Sdept）

系主任姓名（Mname）、课程名（Cname）

成绩（Grade）

单一的关系模式： EA<U、F>

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

©2003-2005，向阳，保留所有权利

数据依赖对关系模式的影响（续）

现实世界的已知事实(语义)告诉我们:

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(正职)负责人;

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 每个学生学习每一门课程有一个成绩。

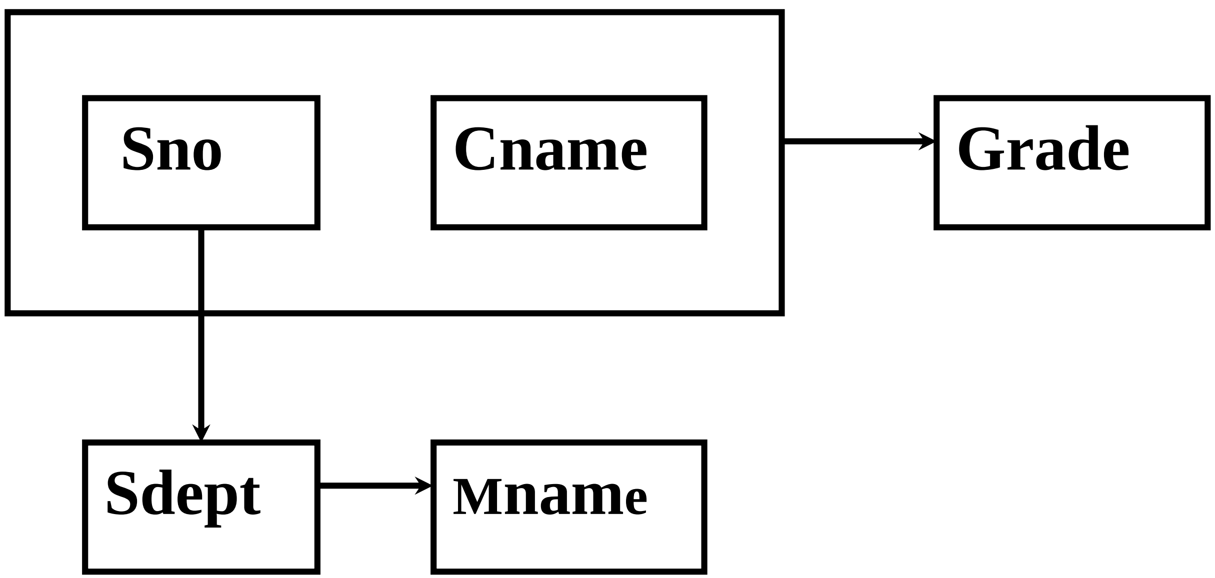
于是得到属性组 U 上的一组函数依赖 F(如图 6.1 所示)。

©2003-2005，向阳，保留所有权利

数据依赖对关系模式的影响（续）

属性组U上的一组函数依赖F：

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



©2003-2005，向阳，保留所有权利

数据依赖对关系模式的影响（续）

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
…	…	…	…	…

©2003-2005，向阳，保留所有权利

关系模式Student<U, F>中存在的问题

1. 数据冗余太大

系、系主任重复太多。

2. 更新异常（Update Anomalies）

换系主任后，每个学生于系主任关系都必须更改。

3. 插入异常（Insertion Anomalies）

如果一个系刚成立，尚无学生，就无法把这个系及其系主任信息插入数据库。

4. 删除异常（Deletion Anomalies）

如果某个系的学生全部毕业，再删除该系学生信息时，把这系及其系主任信息也删掉了。

©2003-2005, 向阳, 保留所有权利

数据依赖对关系模式的影响（续）

结论：

- Student关系模式不是一个好的模式。
- “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

原因：由存在于模式中的某些数据依赖引起的

解决方法：通过分解关系模式来消除其中不合适的数据依赖

分解关系模式

- 把这个单一模式分成3个关系模式：

S（Sno，Sdept，Sno → Sdept）；

SC（Sno，Cno，Grade，（Sno，Cno）→ Grade）；

DEPT（Sdept，Mname，Sdept→ Mname）

©2003-2005，向阳，保留所有权利

分解后的数据表

Sno	Sdept
S1	计算机系
S2	计算机系
S3	计算机系
S4	计算机系
S5	计算机系

表：S

Sno	Cno	Grade
S1	C1	95
S2	C1	90
S3	C1	88
S4	C1	70
S5	C1	78

表：SC

Sdept	Mname
计算机系	张明

表：DEPT

©2003-2005，向阳，保留所有权利

第六章 关系数据理论

©2003-2005，向阳，保留所有权利

6.2 规范化

规范化理论正是用来改造关系模式，通过分解关系模式来消除其中不合适的数据依赖，以解决插入异常、删除异常、更新异常和数据冗余问题。

©2003-2005，向阳，保留所有权利

6.2 规范化

6.2.1 函数依赖

6.2.2 码

6.2.3 范式

6.2.4 2NF

6.2.5 3NF

6.2.6 BCNF

6.2.7 多值依赖

6.2.8 4NF

6.2.9 规范化小结

©2003-2005，向阳，保留所有权利

6.2.1 函数依赖

- 函数依赖
- 平凡函数依赖与非平凡函数依赖
- 完全函数依赖与部分函数依赖
- 传递函数依赖

©2003-2005，向阳，保留所有权利

一、函数依赖

定义6.1 设R(U)是一个属性集U上的关系模式，X和Y是U的子集。

若对于R(U)的任意一个可能的关系r，r中不可能存在两个元组在X上的属性值相等，而在Y上的属性值不等，则称"X函数确定Y"或"Y函数依赖于X"，记作X→Y。

注：1对1和多对1是函数依赖，1对多不是函数依赖

©2003-2005，向阳，保留所有权利

说明

1. 所有关系实例均要满足

2. 语义范畴的概念（学号，性别）（书号，性别）

3. 数据库设计者可以对现实世界作强制的规定

   如：规定姓名不重名

©2003-2005，向阳，保留所有权利

二、平凡函数依赖与非平凡函数依赖

在关系模式R(U)中，对于U的子集X和Y，

如果X→Y，但$Y \nsubseteq X$，则称X→Y是非平凡的函数依赖

若X→Y，但$Y \subseteq X$，则称X→Y是平凡的函数依赖

- 例：在关系SC(Sno, Cno, Grade)中，

  非平凡函数依赖：(Sno, Cno) → Grade

  平凡函数依赖：(Sno, Cno) → Sno

  (Sno, Cno) → Cno

注：平凡函数依赖总是成立的，以后只讨论非平凡的

© 2003-2005，向阳，保留所有权利

平凡函数依赖与非平凡函数依赖（续）

– 若$X \rightarrow Y$，则X称为这个函数依赖的决定属性组，也称为决定因素（Determinant）。

– 若$X \rightarrow Y$，$Y \rightarrow X$，则记作$X \longleftrightarrow Y$。

– 若Y不函数依赖于X，则记作$X \nrightarrow Y$。

©2003-2005，向阳，保留所有权利
同济大学 TONGJI UNIVERSITY

三、完全函数依赖与部分函数依赖

定义6.2 在R(U)中，如果X→Y，并且对于X的任何一个真子集X'，都有X'$\nrightarrow$Y, 则称Y对X完全函数依赖，记作

$X \xrightarrow{F} Y$。

若X→Y，但Y不完全函数依赖于X，则称Y对X部分函数依赖，记作$X \xrightarrow{P} Y$。

注：若X是单属性，则必有$X \xrightarrow{F} Y$；只有当X是属性组时才可能有$X \xrightarrow{P} Y$

©2003-2005，向阳，保留所有权利

完全函数依赖与部分函数依赖（续）

[例1] 中(Sno,Cno)$\xrightarrow{F}$Grade是完全函数依赖，

(Sno,Cno)$\xrightarrow{P}$Sdept是部分函数依赖

因为Sno →Sdept成立，且Sno是（Sno，Cno）的真子集

©2003-2005，向阳，保留所有权利

四、传递函数依赖

定义6.3 在R(U)中，如果X→Y，$(Y \nsubseteq X)$,$Y \nrightarrow X$ Y→Z， 则称Z对X传递函数依赖。

记为：$X \xrightarrow{\text{传递}} Z$

注: 如果Y→X， 即X←→Y，则Z直接依赖于X，表示为：$X \xrightarrow{\text{直接}} Z$，而非传递函数依赖。

例1: 在关系Std(Sno, Sdept, Mname)中，有：

Sno → Sdept， Sdept → Mname

Mname传递函数依赖于Sno

例2: 在关系Std(Sno, Sname,Sdept)中Sname 不重名，是否有Sno →Sname， Sname →Sdept则Sno $\xrightarrow{\text{传递}}$ Sdept？

©2003-2005，向阳，保留所有权利

6.2 规范化

6.2.1 函数依赖

6.2.2 码

6.2.3 范式

6.2.4 2NF

6.2.5 3NF

6.2.6 BCNF

6.2.7 多值依赖

6.2.8 4NF

6.2.9 规范化小结

©2003-2005，向阳，保留所有权利

6.2.2 码

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的侯选码（Candidate Key）。

若候选码多于一个，则选定其中的一个做为主码

（Primary Key）。

例：R（学号，姓名，性别，课号，成绩），R中，学号不是码，课号也不是码，而（学号，课号）是码，

注：K是最少属性的集合，否则$K \xrightarrow{P} U$

© 2003-2005，向阳，保留所有权利

码（续）

- 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（Prime attribute） (回顾：实体完整性定义)
  - 不包含在任何码中的属性称为非主属性（Nonprime attribute）或非码属性（Non-key attribute）
- 全码
  - 整个属性组是码，称为全码（All-key）

©2003-2005，向阳，保留所有权利

码（续）

[例2]

关系模式S(<u>Sno</u>,Sdept,Sage)，单个属性Sno是码，

SC（<u>Sno，Cno</u>，Grade）中，（Sno，Cno）是码

[例3]

关系模式R（P，W，A）

P：演奏者　W：作品　A：听众

一个演奏者可以演奏多个作品

某一作品可被多个演奏者演奏

听众可以欣赏不同演奏者的不同作品

码为(<u>P，W，A</u>)，即All-Key

©2003-2005，向阳，保留所有权利

外部码

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（Foreign key）也称外码。（回顾：参照完整性定义）

- 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

© 2003-2005，向阳，保留所有权利

6.2 规范化

©2003-2005，向阳，保留所有权利

6.2.3 范式

- 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)

©2003-2005，向阳，保留所有权利

6.2.3 范式

- 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

©2003-2005，向阳，保留所有权利

6.2 规范化

© 2003-2005，向阳，保留所有权利

6.2.4 2NF

- 1NF的定义

  如果一个关系模式R的所有属性都是不可分的基本数据项，则R∈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

©2003-2005，向阳，保留所有权利

2NF（续）

[例4] 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

- 函数依赖包括：

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

  $Sno \rightarrow Sdept$

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

  $Sno \rightarrow Sloc$

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

  $Sdept \rightarrow Sloc$

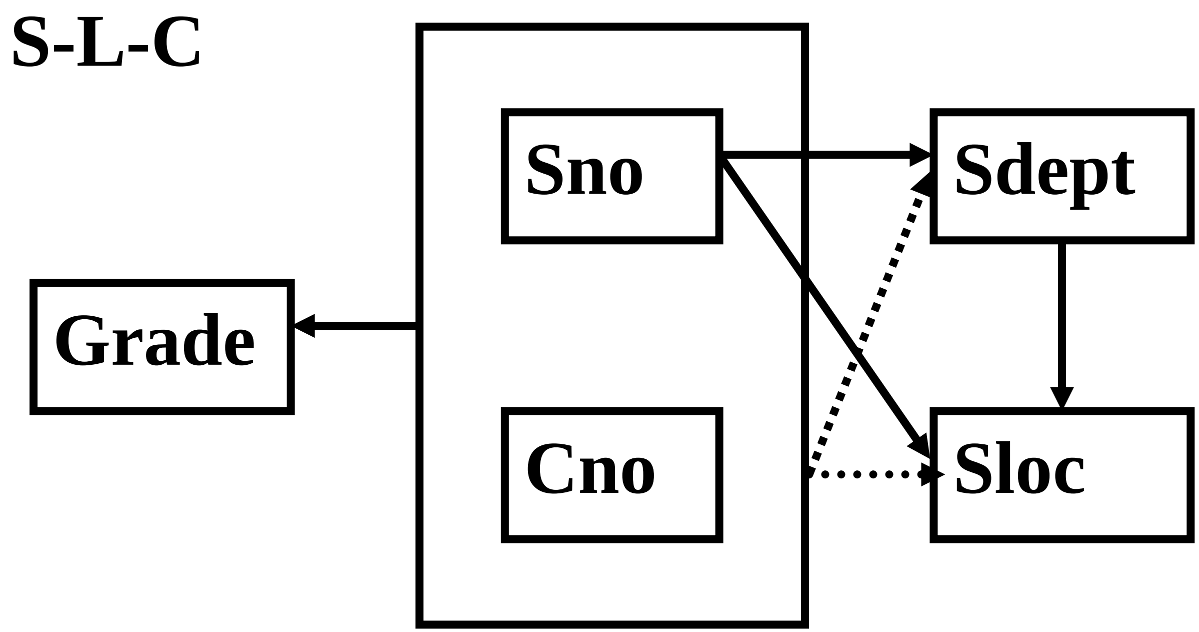
©2003-2005，向阳，保留所有权利

2NF（续）

Sno	Sdept	Sloc	Cno	Grade
S1	计算机系	1号楼	C1	95
S2	计算机系	1号楼	C1	90
S3	计算机系	1号楼	C1	88
S4	计算机系	1号楼	C1	70
S5	计算机系	1号楼	C1	78
…	…	…	…	…

©2003-2005，向阳，保留所有权利

2NF（续）



- 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)

©2003-2005，向阳，保留所有权利

S-L-C不是一个好的关系模式（续）

(1) 数据冗余度大

(2) 插入异常（若插入的学生没选课，则课号没有，同时课号若是主属性则不能为空）

(3) 删除异常（若一个学生开始选了一门课，现又不选了，要删除这门课，由于课号是主属性，要删除，该元组都要删除，则学生信息也删除了）

(4) 修改复杂（若一个学生转到数学系，不但要修改系名，还要修改住处名）

©2003-2005，向阳，保留所有权利

S-L-C不是一个好的关系模式（续）

- 原因

  Sdept、 Sloc部分函数依赖于码。（依赖于Sno）

- 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

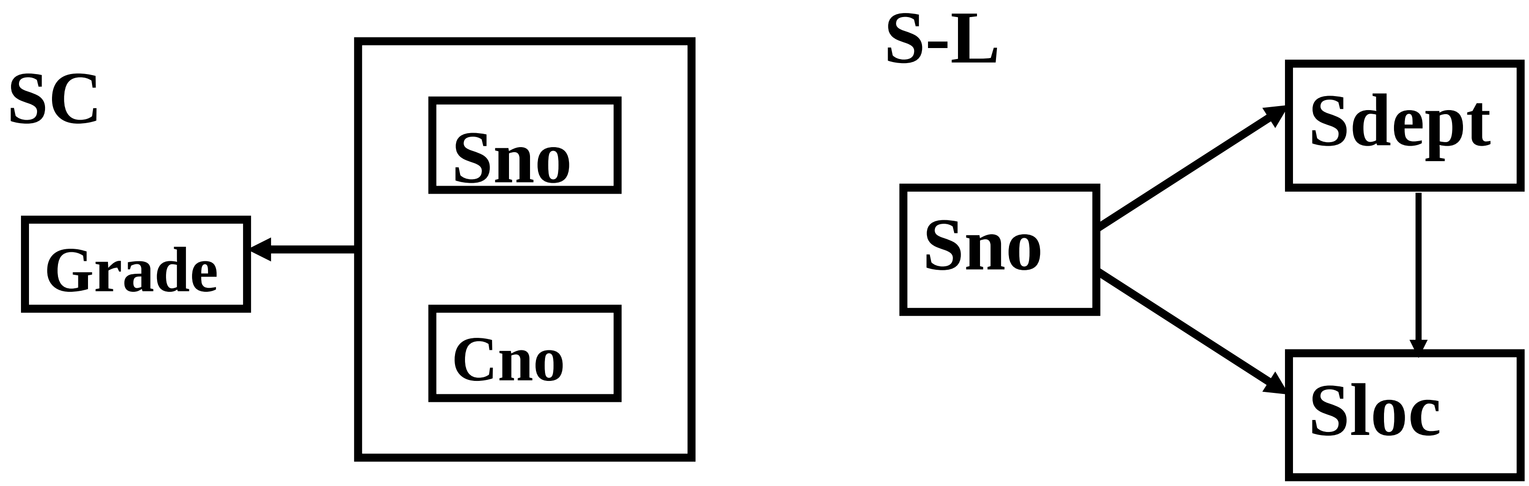
  SC（Sno， Cno， Grade）

  S-L（Sno， Sdept， Sloc）

©2003-2005，向阳，保留所有权利

2NF（续）

函数依赖图：



- 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

©2003-2005，向阳，保留所有权利

2NF（续）

- 2NF的定义

定义6.6 若R∈1NF，且每一个非主属性完全函数依赖于码，则R∈2NF。

例：S-L-C(Sno, Cno, Sdept, Sloc, Grade) ∈1NF

S-L-C(Sno, Cno, Sdept, Sloc, Grade) ∉2NF

SC（Sno， Cno， Grade） ∈ 2NF

S-L（Sno， Sdept， Sloc） ∈ 2NF

（即：消除每一个非主属性对码的部分函数依赖）

©2003-2005，向阳，保留所有权利

同济大学 TONGJI UNIVERSITY

2NF（续）

- 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。

©2003-2005，向阳，保留所有权利

6.2 规范化

6.2.1 函数依赖

6.2.2 码

6.2.3 范式

6.2.4 2NF

6.2.5 3NF

6.2.6 BCNF

6.2.7 多值依赖

6.2.8 4NF

6.2.9 规范化小结

©2003-2005，向阳，保留所有权利

6.2.5 3NF

- 3NF的定义

定义6.7 关系模式$R<U，F>$ 中若不存在这样的码X、属性组Y及非主属性Z（$Z \not\subseteq Y$）, 使得$X \rightarrow Y$，$Y \rightarrow Z$成立，$Y \not\rightarrow X$，则称$R<U，F> \in 3NF$。

■ 若$\boldsymbol{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。

提醒：满足2NF的R中没有非主属性对码的部分依赖

©2003-2005，向阳，保留所有权利

3NF（续）

例：2NF关系模式S-L(Sno, Sdept, Sloc)中

– 函数依赖：

Sno→Sdept

Sdept $\nrightarrow$ Sno

Sdept→Sloc

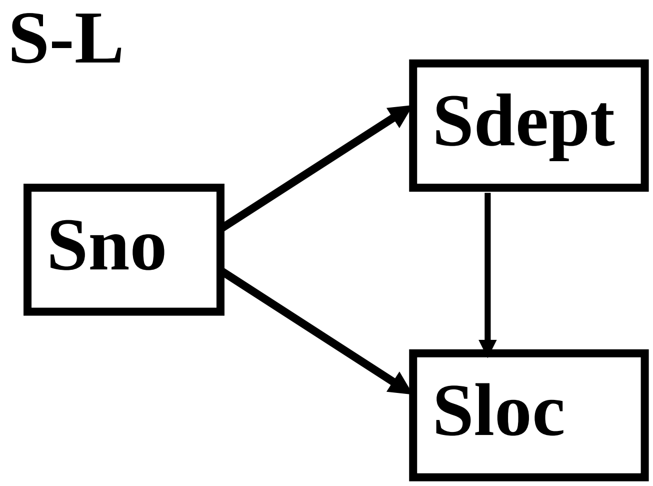
可得：

Sno $\xrightarrow{\text{传递}}$ Sloc，即S-L中存在非主属性对码的传递函数依赖，S-L $\notin$ *3NF*

©2003-2005，向阳，保留所有权利

3NF（续）

函数依赖图：



©2003-2005，向阳，保留所有权利

3NF（续）

- 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）

S-D的码为Sno， D-L的码为Sdept。

- 分解后的关系模式S-D与D-L中不再存在传递依赖

©2003-2005，向阳，保留所有权利

3NF（续）

S-D的码为Sno， D-L的码为Sdept

Sno → **Sdept**　　　　**Sdept** → **Sloc**

S-D　　　　**D-L**

- S-L(Sno, Sdept, Sloc) ∈ 2NF
  S-L(Sno, Sdept, Sloc) ∉ 3NF
  S-D(Sno，Sdept) ∈ 3NF
  D-L(Sdept， Sloc)∈ 3NF

©2003-2005，向阳，保留所有权利
同济大学 TONGJI UNIVERSITY

3NF（续）

- 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。

©2003-2005，向阳，保留所有权利

6.2 规范化

©2003-2005，向阳，保留所有权利

6.2.6 BC范式（BCNF）

- **定义6.8** 关系模式R<U，F>∈1NF，若X→Y且Y $\nsubseteq$ X时,X必含有码，则R<U，F>∈BCNF。（BCNF是修正的3NF）
- 等价于：每一个决定因素都包含码

例：决定因素不包含码的例子：STJ（S，T，J），若规定一个教师只教一门课，则有T →J，T是决定因素但不包含码。

例： R（Sno,Cno,Sname,Grade）,Sname不重名和重名时R是否属于BCNF？R有几个候选码？

©2003-2005，向阳，保留所有权利

BCNF（续）

- 若R∈BCNF，则有结论：
  - 所有非主属性对每一个码都是完全函数依赖（否则就有决定因素不包含码了）
  - 所有的主属性对每一个不包含它的码，也是完全函数依赖。前例sname是主属性，如果对码（sno，cno）有部分函数依赖，则有决定因素不包含码了）
  - 没有任何属性完全函数依赖于非码的任何一组属性（否则就有决定因素不包含码了）

©2003-2005，向阳，保留所有权利

三个结论及证明

- **若R∈BCNF**
- **结论1**：所有非主属性对于每一个码都是完全函数依赖
- **证明**：若某一非主属性Y对某一个码X不完全函数依赖，则码X中的某一主属性Z(真子集)→Y，因为R∈BCNF，即R中的每一个决定因素都是码，则Z是码，这与X是码矛盾。（X是码X的真子集不能是码）
- **结论2**：所有的主属性对每一个不包含它的码，也是完全函数依赖
- **证明**：存在一主属性Z和不包含它的码X，因为X是码$X \rightarrow Z$成立，$X \rightarrow Z$分有两种情况：1)$X \xrightarrow{F} Z$ 2)$X \xrightarrow{P} Z$，若Z不完全函数依赖X(即$X \xrightarrow{P} Z$)，有$X' \subseteq X$ $X' \rightarrow Z$因为R∈BCNF所以X'是码，这与X是码矛盾。因为X是码，所以肯定有$X \rightarrow Z$。或者说，对于不包含Z的码，相对于该码来说Z是非主属性则由结论1可知该结论成立

三个结论及证明（续）

- **结论3：** 没有任何属性完全函数依赖于非码的任何一组属性
- **证明：** 存在一个属性Z完全函数依赖于非码的某一属性组X，即 $X \xrightarrow{F} Z$。因为R∈BCNF所以决定因素都是码，则X是码，这与X是非码矛盾，所以该结论成立。

©2003-2005，向阳，保留所有权利

BCNF（续）

[例5] 关系模式C（<u>Cno</u>，Cname，Pcno）

- C∈3NF
- C∈BCNF

[例6] 关系模式S（<u>Sno</u>，<u>Sname</u>，Sdept，Sage）

- 假定S有两个码Sno，Sname
- S∈3NF。
- S ∈ BCNF

©2003-2005，向阳，保留所有权利

BCNF（续）

［例7］关系模式SJP（S，J，P），语义如下：

S—学生，J—课程，P—名次

1、每个学生选修每门课程的成绩有一定的名次；

2、每门课程中每一名次只有一个学生（没并列）。

- 函数依赖：（S，J）→P；(J，P）→S
- （S，J）与（J，P）都可以作为候选码,属性相交
- SJP∈3NF，
- SJP∈BCNF

©2003-2005，向阳，保留所有权利

BCNF（续）

[例8]在关系模式STJ（S，T，J）中，S表示学生，T表示教师，J表示课程。语义：每一教师只教一门课，每门课有若干教师，某一学生选定某门课后就对应一个固定教师。

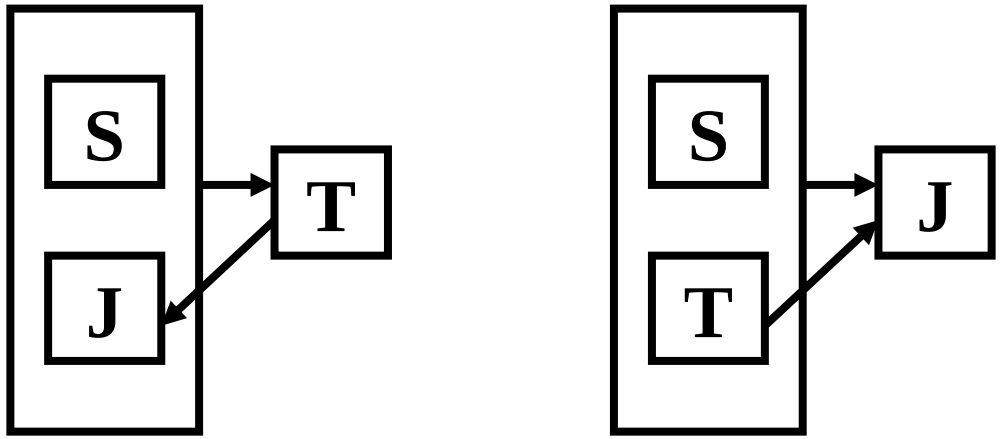
– 函数依赖：

(S，J)→T，(S，T)→J，T→J

– (S，J)和(S，T)都是候选码

©2003-2005，向阳，保留所有权利

BCNF（续）



STJ中的函数依赖

©2003-2005，向阳，保留所有权利

BCNF（续）

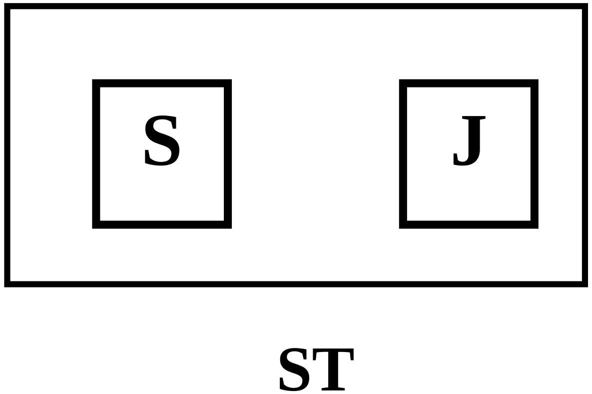
- STJ∈3NF
  - 没有任何非主属性对码传递依赖或部分依赖
- STJ∉BCNF
  - T是决定因素，T不包含码

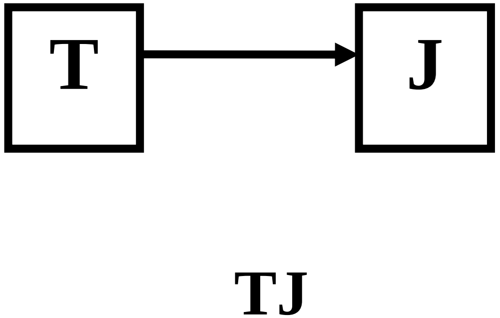
© 2003-2005，向阳，保留所有权利

BCNF（续）

- 解决方法：将STJ分解为二个关系模式：

  ST(S，T) ∈ BCNF， TJ(T，J)∈ BCNF





没有任何属性对码的部分函数依赖和传递函数依赖

©2003-2005，向阳，保留所有权利

3NF与BCNF的关系

- $R \in BCNF \underset{\text{不必要}}{\overset{\text{充分}}{\rightleftarrows}} R \in 3NF$
- 如果R∈3NF，且R只有一个候选码

$R \in BCNF \underset{\text{必要}}{\overset{\text{充分}}{\rightleftarrows}} R \in 3NF$（X->U,Y ⊆U,X->Y,若 Y->Z,X->Z，存在的传递)

注：关于候选码有以下结论：

- 若有属性不在函数依赖集中出现，那么它必须包含在候选码中
- 若有属性不在函数依赖集中任何函数的右边出现，那么它必包含在候选码中
- 若有属性只在函数依赖集的左边出现，则该属性一定包含在候选码中

©2003-2005，向阳，保留所有权利

3NF与BCNF的关系

结论:如果一个关系模式中的所有关系都属于BCNF,那么在函数依赖范畴内,它就实现的彻底的分解,已消除了插入和删除的异常

©2003-2005，向阳，保留所有权利
同济大学 TONGJI UNIVERSITY

6.2 规范化

©2003-2005，向阳，保留所有权利

6.2.7 多值依赖

[例9] 学校中某一门课程由多个教师讲授，他们使用相同的一套参考书。每个教员可以讲授多门课程，每种参考书可以供多门课程使用。

©2003-2005，向阳，保留所有权利

❖ 非规范化关系

课 程 C	教 员 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	数学分析 … …
⋮	⋮	⋮

©2003-2005，向阳，保留所有权利

多值依赖（续）

- 用二维表表示Teaching属于BCNF（下面还有计算数学）

课程C	教员T	参考书B
t物 理	李 勇	普通物理学
w物 理	李 勇	光学原理
物 理	李 勇	物理习题集
v物 理	王 军	普通物理学
s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

… … …

©2003-2005，向阳，保留所有权利

多值依赖（续）

- Teaching∈BCNF
- Teaching具有唯一候选码(C，T，B)， 即全码

BT →→ C？ 、 B ⟶ C？

(数学分析、张平)—>数学

(数学分析、张平)—>计算数学

©2003-2005，向阳，保留所有权利

多值依赖（续）

Teaching模式中存在的问题

(1)数据冗余度大

(2)插入操作复杂

(3) 删除操作复杂

(4) 修改操作复杂

存在
多值依赖

©2003-2005，向阳，保留所有权利

多值依赖（续）

- **定义6.9**

设R(U)是一个属性集U上的一个关系模式， X、 Y和Z是U的子集，并且Z＝U－X－Y。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，给定的一对（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

注：多值依赖中两个属性中的关系还要考察其他属性，即与第三者有关（与其无关也是一种关系），而函数依赖只讨论两个属性，与第三者无关。

例 Teaching（C, T, B）

对于一个（物理，光学原理）有一组T值{李勇、王军}，它仅仅决定于C上的（物理），因为对另一个（物理，普通物理）它对应的一组T值仍是{李勇、王军}，尽管B值已改变。

©2003-2005，向阳，保留所有权利

多值依赖（续）

- 多值依赖的另一个等价的形式化的定义：

对R（U）的任一关系r，如果存在元组t，s 使得$t[X]=s[X]$，那么就必然存在元组 w，$v\in r$，（w，v可以与s，t相同），使得$w[X]=v[X]=t[X]$，而$w[Y]=t[Y]$，$w[Z]=s[Z]$，$v[Y]=s[Y]$，$v[Z]=t[Z]$（即交换s，t元组的Y值所得的两个新元组必在r中），则Y多值依赖于X，记为$X\rightarrow\rightarrow Y$。这里，$X$，$Y$是$U$的子集，$Z=U-X-Y$。

©2003-2005，向阳，保留所有权利

多值依赖（续）

- 平凡多值依赖和非平凡的多值依赖
  - 若$X \to\to Y$，而$Z = \varphi$，则称 $X \to\to Y$为平凡的多值依赖
  - 否则称$X \to\to Y$为非平凡的多值依赖

©2003-2005，向阳，保留所有权利

多值依赖（续）

［例10］关系模式WSC（W，S，C）

- W表示仓库，S表示保管员，C表示商品
- 假设每个仓库有若干个保管员，有若干种商品
- 每个保管员保管所在的仓库的所有商品
- 每种商品被所有保管员保管

©2003-2005，向阳，保留所有权利

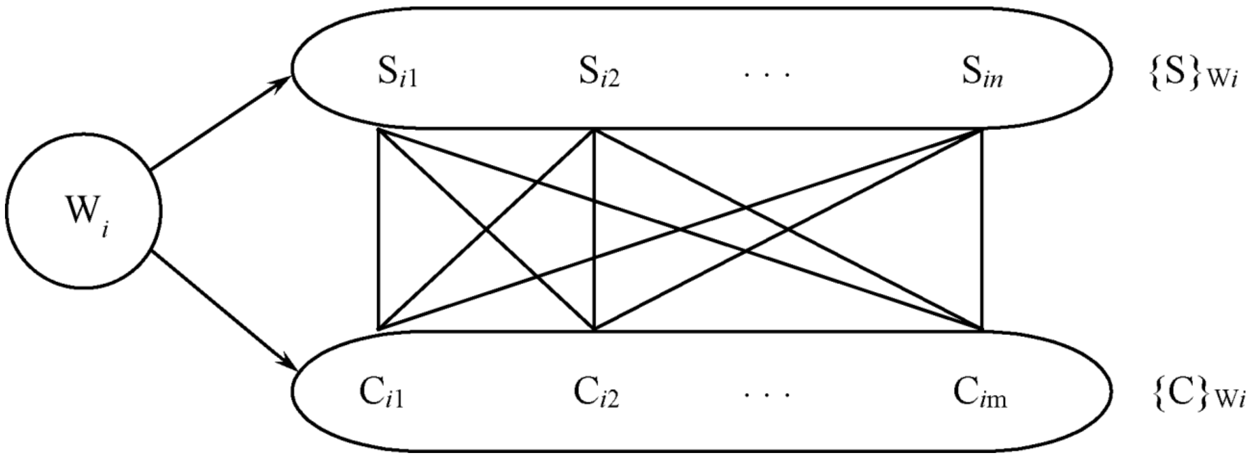
多值依赖（续）

W	S	C
W1	S1	C1
W1	S1	C2
W1	S1	C3
W1	S2	C1
W1	S2	C2
W1	S2	C3
W2	S3	C4
W2	S3	C5
W2	S4	C4
W2	S4	C5

©2003-2005，向阳，保留所有权利

多值依赖（续）

用下图表示这种对应



W→→S且W→→C

©2003-2005，向阳，保留所有权利

多值依赖的性质

（1）多值依赖具有对称性

若X→→Y，则X→→Z，其中Z＝U－X－Y

（2）多值依赖具有传递性

若X→→Y，Y→→Z， 则X→→Z –Y

（3）函数依赖是多值依赖的特殊情况。

若X→Y，则X→→Y。

（4）若X→→Y，X→→Z，则X→→Y∪ Z。

（5）若X→→Y，X→→Z，则X→→Y∩Z。

（6）若X→→Y，X→→Z，则X→→Y-Z，X→→Z -Y。

©2003-2005，向阳，保留所有权利

多值依赖性质证明

- (1)由定义中Z=U-X-Y知道Y=U-X-Z，于是在定义中令Z、Y互换位置立即得到该性质的证明。
- (2)X多值决定Y，令W=U-X-Y，由(1)知道X多值决定W，同时X多值决定X自己。因为X+W=U-Y，而Z-Y一定是U-Y的一部分因此Z-Y被X+W决定。而X多值决定X+W故X决定Z-Y。
- (4)X多值决定Y知道若W=U-X-Y，那么X决定W。而U-X-YZ一定是U-X-Y（=W）的一部分，故X决定U-X-YZ，由(1)X决定YZ。
- (5)X决定Y，Y交Z是Y的一部分，X决定Y交Z。
- (6)X决定Y，Y-Z是Y的一部分，X决定Y-Z。

©2003-2005，向阳，保留所有权利

多值依赖与函数依赖的区别

(1) 多值依赖的有效性与属性集的范围有关:见书P181

若X→→Y在U上成立，则在W（XY⊆W⊆U）上一定成立，反之不然。

例：U=（X，Y，Z1，Z2，Z3）W=（X，Y，Z1，Z2）

若X→→Y在U上成立，则有一个r中两个元组：

x1 y1 z11 z21 z31

x1 y2 z11 z21 z31

交换y1和y2有

x1 y2 z11 z21 z31

x1 y1 z11 z21 z31

它们还在r中。显然，如果不管Z3：

x1 y2 z11 z21

x1 y1 z11 z21

这两个元组也在r中，即X→→Y在U上成立，则在W上也成立。反之不然

©2003-2005，向阳，保留所有权利

多值依赖与函数依赖的区别

(2)

- 若函数依赖X→Y在R（U）上成立，则对于任何Y' ⊂ Y均有X→Y' 成立
- 多值依赖X→→Y若在R(U)上成立，不能断言对于任何Y' ⊂ Y有X→→Y' 成立

©2003-2005，向阳，保留所有权利

6.2 规范化

©2003-2005，向阳，保留所有权利

6.2.8 4NF

- **定义6.10** 关系模式R<U，F>∈1NF，如果对于R的每个非平凡多值依赖X→→Y（Y $\not\subseteq$ X），X都含有码，则R∈4NF。
- 如果R ∈ 4NF， 则R ∈ BCNF
- 4NF限制的是: (见书p181)
  - 不允许有非平凡且非函数依赖的多值依赖(也就是说一个模式中若没有任何非平凡的多值依赖，或者有非平凡的多值依赖但该多值依赖对应的单值依赖是函数的依赖，那么该模式属于4NF)
  - 允许的非平凡的多值依赖是函数依赖

©2003-2005，向阳，保留所有权利

4NF（续）

例： Teaching(C,T,B) $\notin$ 4NF

存在非平凡的多值依赖C→→T，且C不是码

另一说法：因为Teaching(C,T,B) 有非平凡的多值依赖C→→T，但C$\not\to$T是非函数依赖，所以Teaching(C,T,B) $\notin$ 4NF

- 用投影分解法把Teaching分解为如下两个关系模式：

CT(C, T) $\in$ 4NF

CB(C, B) $\in$ 4NF

C→→T， C→→B是平凡多值依赖（因为两个模式中根本就不存在非平凡的多值依赖）

6.2 规范化

6.2.1 函数依赖

6.2.2 码

6.2.3 范式

6.2.4 2NF

6.2.5 3NF

6.2.6 BCNF

6.2.7 多值依赖

6.2.8 4NF

6.2.9 规范化小结

©2003-2005, 向阳, 保留所有权利

6.2.9 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

©2003-2005，向阳，保留所有权利

规范化小结（续）

- 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

消除决定属性集非码的非平凡函数依赖

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

©2003-2005，向阳，保留所有权利

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

©2003-2005，向阳，保留所有权利

规范化小结（续）

- 若只考虑函数依赖，则属于BCNF的关系模式规范化程度是最高的了，若只考虑多值依赖，则属于4NF的关系模式规范化程度是最高的。
- 除函数依赖和多值依赖外还有连接依赖。
- 函数依赖是多值依赖的特例；多值依赖是连接依赖的特例。
- 存在连接依赖4NF的关系模式仍然可能遇到数据冗余、插入、修改和删除异常问题，5NF就是解决这些问题。

©2003-2005，向阳，保留所有权利

例题

- 设有关系模式R(A,B,C,D)及其上的函数依赖集F＝{B→A,BC→D},那么关系模式R最高是哪个范式？
- 设有关系模式R<U, F>，其中U={A, B, C}，F={B→C, A→BC}，则该模式最高满足第几范式？
- 在关系模式中R(A,B,C)中,有函数依赖集F={AB→C,C→B},则R最高达到第几范式？
- 设关系模式R(ABCDEF)的函数依赖及F={A → CD，B → E，AB → F}则R最高属于第几范式？
- 设有关系模式R（U，F），其中：U={A，B，C，D，E}，F={A→D，E→D，D→B，BC→D，DC→A }（1）给出R的候选关键字（2）判断R最高为几级范式？（3）若R不是3NF，将R分解为3NF

©2003-2005，向阳，保留所有权利

例题

- 最后一题答案:
- 1. R的候选码：CE
- 过程：E->D，D->B，所以E->B
- DC->A，所以EC->DC->A
- 由上可得：CE->ABCDE
- 2.R最高为1级范式。
- 理由：首先，R满足属性都不可分，所以是第一范式
- 然后，由EC->D，C->D可知存在部分码依赖，
- 所以，不满足第二范式要求。
- 3.分解为第三范式较复杂（关系比较嵌套），所以分解后的关系比较多:
- {E,D}，E->D
- {D,B}，D->B
- {C,D,A}，CD->A，A->D
- {B,C,D}，BC->D

©2003-2005，向阳，保留所有权利

例题3

- 设关系模式R∈1NF，如果对于R的每个函数依赖X→Y，若Y不是X的子集，则X必含有候选码，则（ ）。

A.R∈1NF

B.R∈2NF

C.R∈3NF

D.R∈4NF

©2003-2005，向阳，保留所有权利

第六章 关系数据理论

6.1 问题的提出

6.2 规范化

6.3 数据依赖的公理系统

***6.4 模式的分解**

6.5 小结

©2003-2005，向阳，保留所有权利

6.3 数据依赖的公理系统

问题提出：

规范化是把低一级的范式通过分解变为高一级的范式，那么分解后的关系模式与原来的关系模式等价吗？如何分解才能保证等价性？

注：等价=属性间语义关系的一致性

© 2003-2005，向阳，保留所有权利

6.3 数据依赖的公理系统

- 逻辑蕴含

定义6.11 对于满足一组函数依赖 *F* 的关系模式*R* <*U*，*F*>，其任何一个关系*r*，若函数依赖$X \rightarrow Y$都成立，（即r中任意两元组*t*，*s*，若t[X]=s[X]，则t[Y]=s[Y]），则称*F*逻辑蕴含$X \rightarrow Y$

例： F={X->Y,Y->Z}关系模式R上的每一个关系r都满足F，且X->Z不属于F,但若X->Z在任何一个r上都成立，则称F逻辑蕴含X->Z，即X->Z隐式包含在F中，即可由F中的依赖导出该依赖函数。

©2003-2005，向阳，保留所有权利

1. Armstrong公理系统

对关系模式$R<U，F>$来说有以下的推理规则：

- A1.自反律（Reflexivity）：若$Y \subseteq X \subseteq U$，则$X \rightarrow Y$为$F$所蕴含。
- A2.增广律（Augmentation）：若$X \rightarrow Y$为F所蕴含，且$Z \subseteq U$，则$XZ \rightarrow YZ$为F所蕴含。
- A3.传递律（Transitivity）：若$X \rightarrow Y$及$Y \rightarrow Z$为F所蕴含，则$X \rightarrow Z$为F所蕴含。

© 2003-2005，向阳，保留所有权利

定理 6.1 Armstrong推理规则是正确的

（l）自反律: 若$Y \subseteq X \subseteq U$，则$X \to Y$为$F$所蕴含

证: 设$Y \subseteq X \subseteq U$

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$Y \subseteq X$，有$t[y]=s[y]$，

所以$X \to Y$成立，自反律得证

© 2003-2005, 向阳，保留所有权利

定理 6.1 Armstrong推理规则是正确的(续）

(2)增广律: 若$X \rightarrow Y$为F所蕴含，且$Z \subseteq U$，则$XZ \rightarrow YZ$ 为F所蕴含。

证：设$X \rightarrow Y$为F所蕴含，且$Z \subseteq U$。

对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$X \rightarrow Y$，于是有$t[Y]=s[Y]$，所以$t[YZ]=s[YZ]$，所以

$XZ \rightarrow YZ$为F所蕴含，增广律得证。

©2003-2005，向阳，保留所有权利

定理 6.1 Armstrong推理规则是正确的（续）

(3) 传递律：若$X \to Y$及$Y \to Z$为F所蕴含，则 $X \to Z$为 F所蕴含。

证：设$X \to Y$及$Y \to Z$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$X \to Y$，有 $t[Y]=s[Y]$；

再由$Y \to Z$，有$t[Z]=s[Z]$，所以$X \to Z$为F所蕴含，传递律得证。

©2003-2005，向阳，保留所有权利

2. 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$X \rightarrow Y$，$X \rightarrow Z$，有$X \rightarrow YZ$。
  （A2， A3）
- 伪传递规则：由$X \rightarrow Y$，$WY \rightarrow Z$，有$XW \rightarrow Z$。
  （A2， A3）
- 分解规则：由$X \rightarrow Y$及 $Z \subseteq Y$，有$X \rightarrow Z$。
  （A1， A3）

© 2003-2005，向阳，保留所有权利

导出规则

2.根据合并规则和分解规则，可得引理6.1

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

©2003-2005，向阳，保留所有权利

3. 函数依赖闭包

定义6.l2 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作 F的闭包，记为F^+。

定义6.13 设F为属性集U上的一组函数依赖，$X \subseteq U$，

$X_F^+ = \{ A | X \rightarrow A$能由$F$ 根据Armstrong公理导出$\}$，

X_F^+称为属性集X关于函数依赖集F 的闭包

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

©2003-2005，向阳，保留所有权利

F的闭包

$F=\{X \rightarrow Y,\ Y \rightarrow Z\}$

$F^{+}=\{$

$X \rightarrow \varphi$, $Y \rightarrow \varphi$, $Z \rightarrow \varphi$, $XY \rightarrow \varphi$, $XZ \rightarrow \varphi$, $YZ \rightarrow \varphi$, $XYZ \rightarrow \varphi$,

$X \rightarrow X$, $Y \rightarrow Y$, $Z \rightarrow Z$, $XY \rightarrow X$, $XZ \rightarrow X$, $YZ \rightarrow Y$, $XYZ \rightarrow X$,

$X \rightarrow Y$, $Y \rightarrow Z$, $XY \rightarrow Y$, $XZ \rightarrow Y$, $YZ \rightarrow Z$, $XYZ \rightarrow Y$,

$X \rightarrow Z$, $Y \rightarrow YZ$, $XY \rightarrow Z$, $XZ \rightarrow Z$, $YZ \rightarrow YZ$, $XYZ \rightarrow Z$,

$X \rightarrow XY$, $XY \rightarrow XY$, $XZ \rightarrow XY$, $XYZ \rightarrow XY$,

$X \rightarrow XZ$, $XY \rightarrow YZ$, $XZ \rightarrow XZ$, $XYZ \rightarrow YZ$,

$X \rightarrow YZ$, $XY \rightarrow XZ$, $XZ \rightarrow XY$, $XYZ \rightarrow XZ$,

$X \rightarrow ZYZ$, $XY \rightarrow XYZ$, $XZ \rightarrow XYZ$, $XYZ \rightarrow XYZ\ \}$

$F=\{X \rightarrow A1, \ldots\ldots, X \rightarrow An\}$的闭包$F^{+}$计算（求出）是一个NP完全问题

© 2003-2005, 向阳，保留所有权利

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，$Y \subseteq U$，$X \rightarrow Y$能由F 根据Armstrong公理导出的充分必要条件是$Y \subseteq X_F^+$

- 用途

  将判定$X \rightarrow Y$是否能由F根据Armstrong公理导出的问题，转化为求出X_F^+、判定Y是否为X_F^+的子集的问题

©2003-2005，向阳，保留所有权利

求闭包的算法

算法6.1 求属性集X（$X \subseteq U$）关于U上的函数依赖集F 的闭包X_F^+

输入：X，F　　　　　　输出：X_F^+

步骤：

（1）令$X^{(0)}=X$，$i=0$

（2）求B，这里$B = \{ A \mid (\exists V)(\exists W)(V \rightarrow W \in F \wedge V \subseteq X^{(i)} \wedge A \in W) \}$；

（3）$X^{(i+1)}=B \cup X^{(i)}$

（4）判断$X^{(i+1)} = X^{(i)}$吗?

（5）若相等或$X^{(i)}=U$，则$X^{(i)}$就是X_F^+，算法终止。

（6）若否，则 $i=i+1$，返回第（2）步。

函数依赖闭包

[例1] 已知关系模式$R<U，F>$，其中

$U=\{A，B，C，D，E\}$；

$F=\{AB\rightarrow C，B\rightarrow D，C\rightarrow E，EC\rightarrow B，AC\rightarrow B\}$。

求（AB）$_F^+$。

解 设$X^{(0)}=AB$;

(1) $X^{(1)}=AB\cup CD=ABCD$。

(2) $X^{(0)}\neq X^{(1)}$

$X^{(2)}=X^{(1)}\cup BE=ABCDE$。

(3) $X^{(2)}=U$，算法终止

→（AB）$_F^+=ABCDE$。 见书P185

©2003-2005，向阳，保留所有权利

设 $X^{(0)}=AB$

计算 $X^{(1)}$；逐一的扫描 F 集合中各个函数依赖，找左部为 A，B 或 AB 的函数依赖。得到两个：$AB\to C$，$B\to D$。于是 $X^{(1)}=AB\cup CD=ABCD$。

$B\to D$, $AB\to C$

因为 $X^{(0)}\neq X^{(1)}$，所以再找出左部为 $ABCD$ 子集的那些函数依赖，又得到 $C\to E$，$AC\to B$，于是 $X^{(2)}=X^{(1)}\cup BE=ABCDE$。

{A, B, C, D, AB, AC, AD, BC, BD, CD, ABC, ABD, BCD, ABCD}

因为 $X^{(2)}$ 已等于全部属性集合，所以 $(AB)_F^+=ABCDE$。

同濟大學 TONGJI UNIVERSITY

算法6.1

对于算法6.1， 令$a_i=|X^{(i)}|$，$\{a_i\}$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。

对上例：$|X^{(0)}|=|AB|=2$, $|X^{(1)}|=|ABCD|=4$,$|X^{(2)}|=|ABCDE|=5$

©2003-2005，向阳，保留所有权利

4. Armstrong公理系统的有效性与完备性

- 定理6.2 Armstrong公理系统是有效的、完备的
- 证明：

1. 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）

   可由定理6.1得证

2. 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）

   只需证明逆否命题: 若函数依赖$X \rightarrow Y$不能由*F*从Armstrong公理导出，那么它必然不为*F*所蕴含

Armstrong公理系统完备性证明

(1) 引理: 若$V \to W$成立，且$V \subseteq X_F^+$，则$W \subseteq X_F^+$

(2) 构造一张二维表r，它由下列两个元组构成，可以证明r必是R（U，F）的一个关系，即F^+中的全部函数依赖在 r上成立。

$$\begin{array}{cc} \overbrace{}^{X_F^+} & \overbrace{}^{U-X_F^+} \\ 11......1 & 00......0 \\ 11......1 & 11......1 \end{array}$$

(3) 若$X \to Y$ 不能由F从Armstrong公理导出，则Y 不是X_F^+的子集。

©2003-2005，向阳，保留所有权利

定理 6.2 Armstrong 公理系统是有效的、完备的。

Armstrong 公理系统的有效性可由**定理 6.1** 得到证明。这里给出完备性的证明。

证明完备性的逆否命题，即若函数依赖 $X \to Y$ 不能由 F 从 Armstrong 公理导出，那么它必然不为 F 所蕴含，它的证明分三步。

(1) 若 $V \to W$ 成立，且 $V \subseteq X_F^+$，则 $W \subseteq X_F^+$。

证 因为 $V \subseteq X_F^+$，所以有 $X \to V$ 成立；于是 $X \to W$ 成立（因为 $X \to V, V \to W$），所以 $W \subseteq X_F^+$。

(2) 构造一张二维表 r，它由下列两个元组构成，可以证明 r 必是 $\boldsymbol{R}(U,F)$ 的一个关系，即 F 中的全部函数依赖在 r 上成立。

$$\begin{array}{cc} \overbrace{\quad\quad\quad}^{X_F^+} & \overbrace{\quad\quad\quad}^{U - X_F^+} \\ 11\cdots\cdots1 & 00\cdots\cdots0 \\ 11\cdots\cdots1 & 11\cdots\cdots1 \end{array}$$

若 r 不是 $\boldsymbol{R}<U,F>$ 的关系，则必由于 F 中有某一个函数依赖 $V \to W$ 在 r 上不成立所致。由 r 的构成可知，V 必定是 X_F^+ 的子集，而 W 不是 X_F^+ 的子集，可是由第(1)步，$W \subseteq X_F^+$，矛盾。所以 r 必是 $\boldsymbol{R}<U,F>$ 的一个关系。

(3) 若 $X \to Y$ 不能由 F 从 Armstrong 公理导出，则 Y 不是 X_F^+ 的子集，因此必有 Y 的子集 Y' 满足 $Y' \subseteq U - X_F^+$，则 $X \to Y$ 在 r 中不成立，即 $X \to Y$ 必不为 $\boldsymbol{R}<U,F>$ 蕴含。

Armstrong 公理的完备性及有效性说明了“导出”与“蕴含”是两个完全等价的概念。于是 F^+ 也可以说成是由 F 出发借助 Armstrong 公理导出的函数依赖的集合。

从蕴含（或导出）的概念出发，又引出了两个函数依赖集等价和最小依赖集的概念。

©2003-2005，向阳，保留所有权利

5. 函数依赖集等价

定义6.14 如果$G^+=F^+$，就说函数依赖集F覆盖G（F是G的覆盖，或G是F的覆盖），或F与G等价。

引理6.3 $F^+ = G^+$的充分必要条件是$F \subseteq G^+$，和$G \subseteq F^+$

证: 必要性显然，只证充分性。

（1）若$F \subseteq G^+$，则$X_F{}^+ \subseteq X_{G+}{}^+$。

（2）任取$X \rightarrow Y \in F^+$ 则有 $Y \subseteq X_F{}^+ \subseteq X_{G+}{}^+$。

所以$X \rightarrow Y \in (G^+)^+ = G^+$。即$F^+ \subseteq G^+$。

（3）同理可证$G^+ \subseteq F^+$，所以$F^+ = G^+$。

© 2003-2005，向阳，保留所有权利

5. 函数依赖集等价

又一证法: 必要性显然，只证充分性。

（1）因为$F \subseteq G^+$，则F可由从G出发导出，所以F^+可由从G出发导出，而从G出发导出的函数依赖的全部就是G^+

所以$F^+ \subseteq G^+$。

（2）同理可证$G^+ \subseteq F^+$，

（3）所以$F^+ = G^+$。

©2003-2005，向阳，保留所有权利

6. 最小依赖集

定义6.15 如果函数依赖集F满足下列条件，则称F为一个极小函数依赖集。亦称为最小依赖集或最小覆盖。

(1) F中任一函数依赖的右部仅含有一个属性。

(2) F中不存在这样的函数依赖X→A，使得F与F-{X→A}等价。

(3) F中不存在这样的函数依赖X→A， X有真子集Z使得F-{X→A}∪{Z→A}与F等价。

©2003-2005, 向阳，保留所有权利

最小依赖集

[例2] 关系模式*S*<*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

©2003-2005，向阳，保留所有权利

7. 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F_m。此F_m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。

©2003-2005，向阳，保留所有权利

极小化过程（续）

(1)逐一检查F中各函数依赖FD_i：$X\rightarrow Y$，若$Y=A_1A_2\ldots A_k$，$k>2$，则用 { $X\rightarrow A_j$ |j=1，2，…，k} 来取代$X\rightarrow Y$。

(2)逐一检查F中各函数依赖FD_i：$X\rightarrow A$，令$G=F-\{X\rightarrow A\}$，若$A\in X_G^+$，则从F中去掉此函数依赖。（注：$A\in X_G^+$说明$X\rightarrow A$ *不在G中显式存在，但能被导出*）（F与G等价的充分必要条件是$A\in X_G^+$）

(3)逐一取出F中各函数依赖FD_i：$X\rightarrow A$，设$X=B_1B_2\ldots B_m$，逐一考查B_i（i=l，2，…，m），若$A\in(X-B_i)_F^+$，则以$X-B_i$ 取代X。*（注：$A\in(X-B_i)_F^+$说明$X-B_i\rightarrow A$可导出$X\rightarrow A$，所以$X-B_i\rightarrow A$可取代$X\rightarrow A$）*

©2003-2005，向阳，保留所有权利

极小化过程（续）

[例3] $F = \{A \to B, B \to A, B \to C, A \to C, C \to A\}$

F_{m1}、F_{m2}都是F的最小依赖集：

$F_{m1} = \{A \to B, B \to C, C \to A\}$

$F_{m2} = \{A \to B, B \to A, A \to C, C \to A\}$

- F的最小依赖集F_m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

©2003-2005，向阳，保留所有权利

极小化过程（续）

- [例4] 3. 已知函数依赖集F={A→B, ABCD→E, EF→G, EF→H, ACDF→EG}，则F的最小覆盖Fm为：
- A. Fm={A→B, ABCD→E, EF→G, EF→H, ACDF→E, ACDF→G }
- B. Fm={A→B, ABCD→E, EF→G, EF→H, ACDF→EG}
- C. Fm={A→B, ACD→E, EF→G, EF→H}
- D. Fm={A→B, ACD→E, EF→H, ACDF→G }

©2003-2005，向阳，保留所有权利

极小化过程（续）

- [例4] 3. 已知函数依赖集F={A→B, ABCD→E, EF→G, EF→H, ACDF→EG}，则F的最小覆盖Fm为：
- A. Fm={A→B, ABCD→E, EF→G, EF→H, ACDF→E, ACDF→G }（违反2）
- B. Fm={A→B, ABCD→E, EF→G, EF→H, ACDF→EG}（违反1）
- C. Fm={A→B, ACD→E, EF→G, EF→H}
- D. Fm={A→B, ACD→E, EF→H, ACDF→G }

©2003-2005，向阳，保留所有权利

极小化过程（续）

结论：两个关系模式R（U，F），S（U，G），如果F与G等价，那么R的关系一定是S的关系，S的关系也一定是R的关系。所以在R（U，F）中用与F等价的依赖集G来取代F是允许的。

©2003-2005，向阳，保留所有权利

第六章 关系数据理论

6.1 问题的提出

6.2 规范化

6.3 数据依赖的公理系统

***6.4 模式的分解**

6.5 小结

©2003-2005, 向阳, 保留所有权利

6.4 模式的分解

- 把低一级的关系模式分解为若干个高一级的关系模式的方法不是唯一的
- 只有能够保证分解后的关系模式与原关系模式等价，分解方法才有意义

©2003-2005，向阳，保留所有权利

关系模式分解的标准

三种模式分解等价的定义：

1. 分解具有无损连接性

2. 分解要保持函数依赖

3. 分解既要保持函数依赖，又要具有无损连接性

©2003-2005，向阳，保留所有权利

模式的分解（续）

定义6.16 关系模式R<U,F>的一个分解：

ρ={ R_1<U_1,F_1>，R_2<U_2,F_2>，…，R_n<U_n,F_n>}

$U=\bigcup_{i=1}^{n} U_i$，且不存在 $U_i \subseteq U_j$，F_i 为 F在 U_i 上的投影

定义6.17 函数依赖集合$\{X\to Y \mid X\to Y \in F^+ \wedge XY\subseteq U_i\}$ 的一个覆盖 F_i 叫作 F 在属性 U_i 上的投影

©2003-2005，向阳，保留所有权利

模式的分解（续）

例：S-L（Sno， Sdept， Mname）

F={ Sno→Sdept,Sdept→ Mname}

S-L∈2NF

分解方法可以有多种：

1. S-L分解为三个关系模式：SN(Sno)
   SD(Sdept)
   SM(Mname)
2. S-L分解为二个关系模式：SS(Sno, Sdept)
   SM (Sno, Mname)
3. S-L分解为二个关系模式：SM(Sno, Sdept)
   DN(Sdept, Mname)

© 2003-2005，向阳，保留所有权利

具有无损连接性的模式分解

- 关系模式R<U,F>的一个分解 ρ={ R_1<U_1,F_1>，R_2<U_2,F_2>， …， R_n<U_n,F_n>}

  若R与R1、R2、…、Rn自然连接的结果相等，则称关系模式R的这个分解ρ具有无损连接性（Lossless join）

- 具有无损连接性的分解保证不丢失信息（对第一种分解作连接就是笛卡儿积，这个笛卡儿积对于查询是无损的，即查询结果没有丢失信息，但有失真 ）
- 无损连接性不一定能解决插入异常、删除异常、修改复杂、数据冗余等问题

©2003-2005，向阳，保留所有权利

模式的分解（续）

○第2、3种分解方法具有无损连接性

但第2种分解方法没有保持原关系中的函数依赖

Sdept→ Mname ，所以导致第二种方法有插入、删除异常，（但信息不丢失，也不失真）

○第3种由于保持了原关系中的函数依赖，所以他解决更新异常，又没有丢失元数据库的信息

©2003-2005，向阳，保留所有权利

保持函数依赖的模式分解

设关系模式R<U,F>被分解为若干个关系模式

R_1<U_1,F_1>，R_2<U_2,F_2>，…，R_n<U_n,F_n>

（其中$U=U_1 \cup U_2 \cup \ldots \cup U_n$，且不存在$U_i \subseteq U_j$，$F_i$为F在$U_i$上的投影），若F所逻辑蕴含的函数依赖一定也由分解得到的某个关系模式中的函数依赖F_i所逻辑蕴含，则称关系模式R的这个分解是保持函数依赖的（Preserve dependency）

© 2003-2005，向阳，保留所有权利

模式的分解（续）

- 如果一个分解具有无损连接性，则它能够保证不丢失信息
- 如果一个分解保持了函数依赖，则它可以减轻或解决各种异常情况
- 分解具有无损连接性和分解保持函数依赖是两个互相独立的标准。具有无损连接性的分解不一定能够保持函数依赖；同样，保持函数依赖的分解也不一定具有无损连接性。

©2003-2005，向阳，保留所有权利

模式的分解（续）

第1种分解方法既不具有无损连接性，也未保持函数依赖，

它不是原关系模式的一个等价分解

第2种分解方法具有无损连接性，但未保持函数依赖

第3种分解方法既具有无损连接性，又保持了函数依赖

©2003-2005，向阳，保留所有权利

分解算法

- **算法6.2** 判别一个分解的无损连接性
- **算法6.3**（合成法）转换为3NF的保持函数依赖的分解。
- **算法6.4** 转换为3NF既有无损连接性又保持函数依赖的分解
- **算法6.5**（分解法）转换为BCNF的无损连接分解
- **算法6.6** 达到4NF的具有无损连接性的分解

©2003-2005，向阳，保留所有权利

第六章 关系数据理论

6.1 问题的提出

6.2 规范化

6.3 数据依赖的公理系统

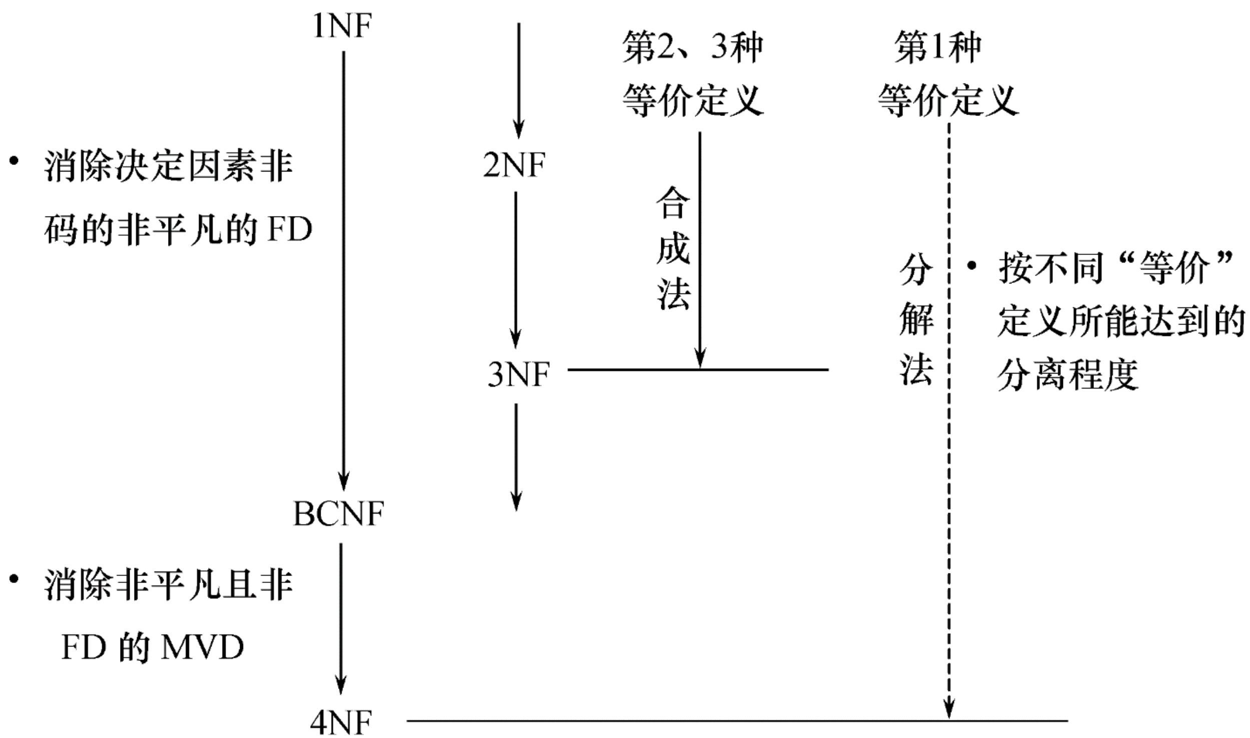
***6.4 模式的分解**

6.5 小结

©2003-2005，向阳，保留所有权利

6.5 小结

关系模式的规范化，其基本思想：



©2003-2005，向阳，保留所有权利

小结(续)

- 若要求分解具有无损连接性，那么模式分解一定能够达到4NF
- 若要求分解保持函数依赖，那么模式分解一定能够达到3NF，但不一定能够达到BCNF
- 若要求分解既具有无损连接性，又保持函数依赖，则模式分解一定能够达到3NF，但不一定能够达到BCNF

©2003-2005，向阳，保留所有权利

小结(续)

- 规范化理论为数据库设计提供了理论的指南和工具
  - 也仅仅是指南和工具
- 并不是规范化程度越高，模式就越好
  - 必须结合应用环境和现实世界的具体情况合理地选择数据库模式

©2003-2005, 向阳, 保留所有权利

下课了。。。

研究

休息一会儿。。。

©2003-2005，向阳，保留所有权利