第六章 关系数据理论

6.1 问题的提出

6.2 规范化

第六章 关系数据理论

6.1 问题的提出

6.2 规范化

6.1 问题的提出

- 数据库的一般概念
  - 关系数据库的基本概念
  - 关系模型的三个部分
    - 关系数据结构；关系的完整性；关系操作
  - 关系数据库的标准语言
- 未涉及的问题
  - 如何构造数据模式
    - 构造几个关系
    - 关系的属性组成
  - 模式的构造是关系数据库的逻辑设计问题

6.1 问题的提出

- 关系模式中属性的相互关连

现实世界的许多已有事实限定了关系模式所有可能的关系必须满足一定的完整性约束条件。这些约束或者通过对属性取值范围的限定，或者通过属性间的相互关连（主要体现于值的相等与否）反映出来。后者称为数据依赖，它是数据模式设计的关键。关系模式应当刻划这些完整性约束条件，于是一个关系模式应当是一个五元组

R（U，D，dom，F）

本章只研究三元组 **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。

6.1 问题的提出

- 第一范式

关系，作为一张二维表，对它有一个最起码的要求：每一个分量必须是不可分的数据项。满足了这个条件的关系模式就属于第一范式

6.1 问题的提出

- 数据依赖

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。它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。

- 两种最重要的数据依赖
  - 函数依赖

    Functional Dependency，简称FD
  - 多值依赖

    Multivalued Dependedcy，简称MVD

6.1 问题的提出

- 函数依赖现象

例如描述一个学生的关系

Student（SNO，SNAME，SDEPT）

在“学号”确定以后，姓名和学生所在的系就确定了。

如同自变量 x 确定之后，相应的函数值 f(x) 也就确定了。称SNO函数决定SNAME和SDEPT，或者说SNAME，SDEPT函数依赖于SNO，记为：

SNO → SNAME

SNO → SDEPT

6.1 问题的提出

- 建立关系模式

现有如下对象要建立关系模式：

学生（用学号Sno描述）

系（用系名Sdept描述）

系负责人（用其姓名Mname描述）

课程（用课程号Cno描述）

成绩（用Grade描述）

于是得到关系模式及属性

U＝{Sno，Sdept，Mname，Cno，Grade}

6.1 问题的提出

- 上例的函数依赖
  - 由现实世界的已知事实（语义）可以推知

（1）一个系有若干学生，但一个学生只属于一个系。

（2）一个系只有一名系主任。

（3）一个学生可以选修多门课程，每门课程有若干学生选修。

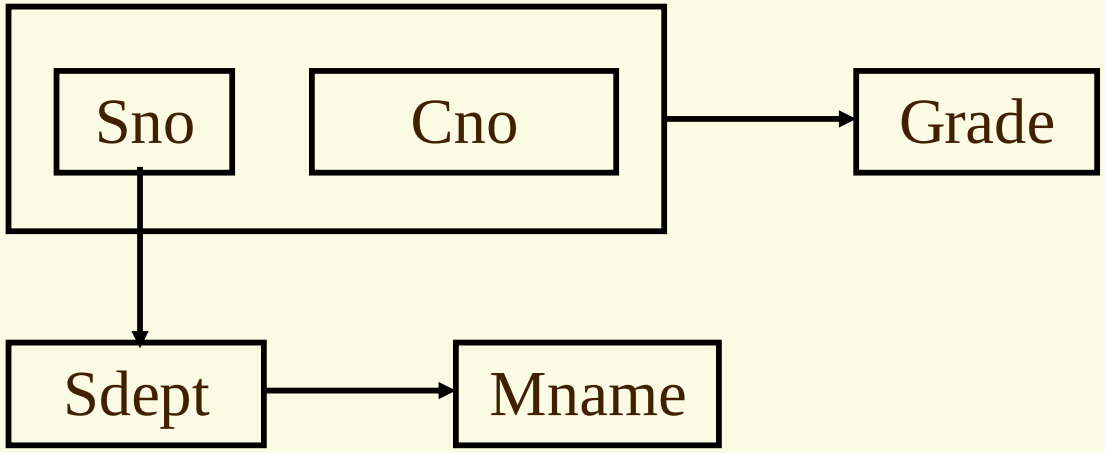
（4）每个学生所学的每门课程都有一个成绩。

6.1 问题的提出

- 上例的函数依赖

根据现实情况得到函数依赖

F＝{Sno→Sdept，Sdept→Mname，（Sno，Cno）→Grade}



6.1 问题的提出

现只考虑函数依赖这一种数据依赖，得到一个描述学生的关系模式：Student(U , F)

假设存在下表的某一时刻Student的一个实例。

Sno	Sdept	Mname	Cname	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
…	…	…	…	…

6.1 问题的提出

- 上例模式属于第一范式，但存在问题
  - 数据冗余太大
    - 每一个系负责人的姓名要与该系学生的每一门功课的成绩出现的次数一样多。
  - 更新异常
    - 由于数据冗余，更新数据时，系统要付出很大的代价来维护数据库的完整性。

6.1 问题的提出

- 插入异常
  - 如果一个系刚成立尚无学生，那么就无法把这个系及其负责人的信息存入数据库。
- 删除异常
  - 如果某个系的学生全部毕业了，在删除该系学生选修课程的同时，把这个系及其负责人的信息也丢掉了。

6.1 问题的提出

- 异常原因

  模式中的函数依赖存在不好的性质
- 解决方法

  将上述模式分解

  S（Sno，Sdept，Sno→Sdept）

  SG（Sno，Cno，Grade，（Sno，Cno）→ Grade）

  DEPT（Sdept，Mname，Sdept → Mname）

第六章 关系数据理论

6.1 问题的提出

6.2 规范化

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

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

6.2.1 函数依赖

- 定义6.1 设R（U）是属性集 U 上的关系模式。X，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X，记作X → Y。
- 注意，函数依赖不是指关系模式 R 的某个或某些关系满足的约束条件，而是指 R 的一切关系均要满足的约束条件。

6.2.1 函数依赖

- 术语和记号
  - X → Y，但 Y⊈ X，则称X → Y是 非平凡的函数依赖。
  - X → Y，但 Y⊆ X，则称X → Y是 平凡的函数依赖。
  - 若 X → Y，则 X 叫做决定因素
  - 若 X → Y，Y → X，则记作X ← →Y。
  - 若 Y 不函数依赖于 X，则记作X↛ Y

6.2.1 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'↛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：$X \xrightarrow{P} Y$

例如：在关系S（SNO，SNAME，SDEPT，SAGE）中，SNO →SDEPT，SNO →SAGE，SNO← →SNAME（无重名）

关系SC（SNO，CNO，G），SNO↛ CNO，SNO↛ G。但（SNO，CNO）$\xrightarrow{F}$ G，（SNO，CNO）就是决定因素。

6.2.1 函数依赖

- 定义6.3 在R（U）中，如果 X → Y，（$Y \not\subseteq X$），$Y \not\to X$，Y → Z，则称 Z 对 X 传递函数依赖。

  加上条件$Y \not\to X$，是因为如果 Y → X，则X ← →Y，实际上是 $X \xrightarrow{直接} Z$，是直接函数依赖而不是传递函数依赖。

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

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为R的候选码（Candidate）。若候选码多于一个，则选定其中的一个为主码（Primary）。

包含在任何一个候选码中的属性，叫做主属性（Prime attribute）。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）。整个属性组是码，称为全码（All-key）

6.2.2 码

- 例如：在关系模式S（SNO，SDEPT，SAGE）中，SNO是码；在关系模式SC（SNO，CNO，G）中属性组合（SNO，CNO）是码。

关系模式R（P，W，A），属性P表示演奏者，W表示作品，A表示听众。该关系模式的码为（P，W，A），即All-key

6.2.2 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。
- 例如：在SC（SNO，CNO，G）中，SNO不是码，但是关系模式S（SNO，SDEPT，SAGE）的码，则SNO是关系模式SC的外部码。
- 主码与外部码提供了一个表示关系间联系的手段。

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

6.2.3 范式

- 关系数据库中的关系要满足一定的要求，满足不同程度要求的为不同范式。所谓“第几范式”是表示关系的某一个级别。所以经常称某一关系范式R为第几范式，记作：R∈xNF。
- 目前范式的分类为：
  - 5NF，4NF，BCNF，3NF，2NF，1NF
  - 上述的范式级别逐步降低
- 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。

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

6.2.4 2NF

- 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。
- 不是2NF的示例

关系模式S-L-C（SNO，SDEPT，SLOC，CNO，G）

其中SLOC为学生住所，每个系的同学住在同一个地方。

码为（SNO，CNO）

6.2.2 2NF

函数依赖有：

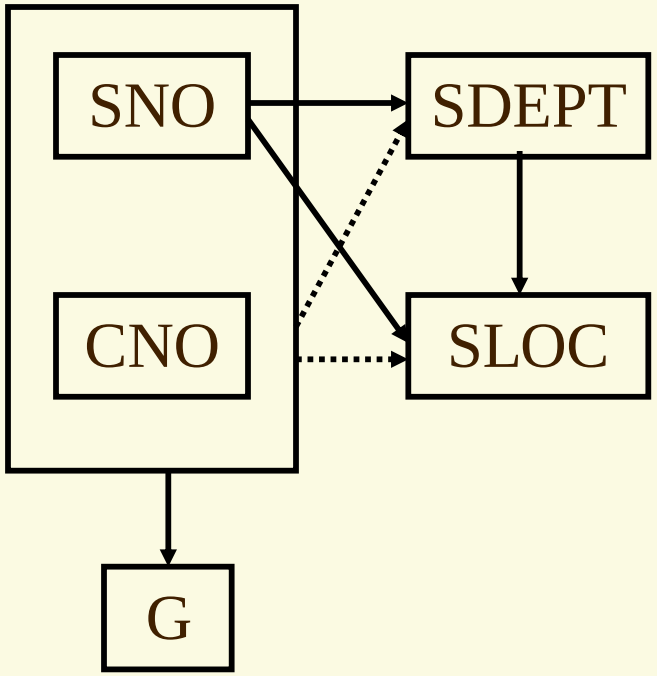
（SNO，CNO）$\xrightarrow{F}$ G

SNO→SDEPT

（SNO，CNO）$\xrightarrow{P}$ SDEPT

SNO →SLOC

（SNO，CNO）$\xrightarrow{P}$ SLOC

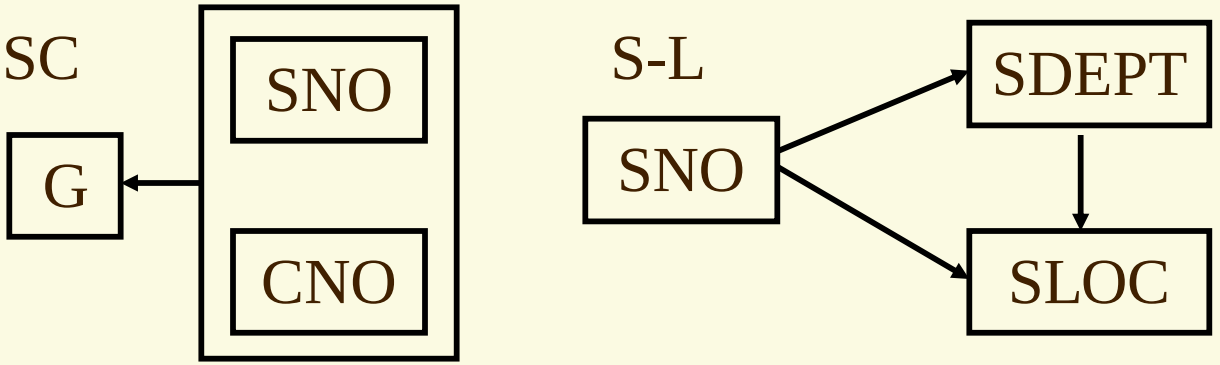


6.2.2 2NF

- 示例分析
  - 非主属性SDEPT，SLOC并不完全函数依赖于码。因此关系模式S-L-C不符合2NF定义。即S-L-C $\notin$ 2NF。
  - 不属于2NF的范式，会产生以下几个问题：
    - 插入异常
    - 删除异常
    - 冗余度大、修改复杂
  - 产生问题的原因是非主属性SDEPT，SLOC对码的不完全依赖。

6.2.2 2NF

- 解决的方法是将关系模式 S-L-C 分解为两个关系模式
  - SC（SNO，CNO，G）
  - S-L（SNO，SDEPT，SLOC）
- 分解后的函数依赖关系如图所示，这样两个关系模式都是2NF范式。



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

6.2.5 3NF

- 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 及非主属性组 Z（Z $\subsetneq$ Y）使得X→Y，（Y↛ X）Y→Z 成立，则称 R（U，F）∈3NF
- 3NF范式中，每一个非主属性既不部分依赖于码也不传递依赖于码。上节中的关系模式S-L不是3NF。
- S-L同样要进行分解，分解后不存在传递依赖
  - S-D（SNO，SDEPT）
  - D-L（SDEPT，SLOC）

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

6.2.6 BCNF

- **BCNF比上述的3NF又进了一步，通常认为是修正的第三范式，有时也称为扩充的第三范式。**
- **定义6.8 关系模式R（U，F）∈1NF。若X → Y且Y ⊈X时，X必含有码，则R（U，F）∈BCNF。**
- **在关系模式R（U，F）中，若每一个决定因素都包含码，则R（U，F）∈BCNF。**

6.2.6 BCNF

- **BCNF关系模式有如下结论：**
  - 所有非主属性对每一个码都是完全函数依赖。
  - 所有的主属性对每一个不包含它的码，也是完全函数依赖。
  - 没有任何属性完全函数依赖于非码的任何一组属性。
- **BCNF一定是3NF，但3NF未必是BCNF。**

6.2.6 BCNF

- 例1
  - 关系模式SJP（S，J，P）中，S是学生，J表示课程，P表示名次。可得如下函数依赖：
    （S，J）→P；（J，P）→S
  - 关系模式中没有属性对码的传递依赖或部分依赖，所以SJP∈3NF
  - 决定因素（S，J）与（J，P）就是码，所以SJP ∈BCNF。

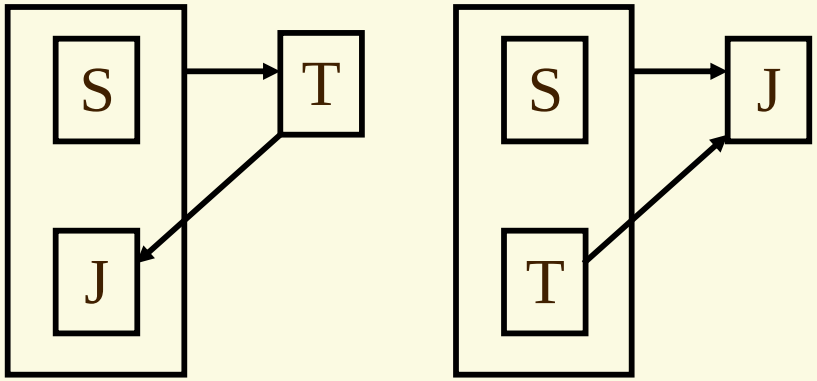
6.2.6 BCNF

- 例2
  - 关系模式STJ（S，T，J）中，S表示学生，T表示教师，J表示课程。假设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，就对应一个固定的教师。则可得如下函数依赖

（S，J）→ T

（S，T）→ J

T → J



6.2.6 BCNF

- （S，J），（S，T）都是候选码，STJ 是 3NF。但STJ 不是 BCNF，因为 T 是决定因素，但 T 不包含码。
- 非 BCNF 关系模式，也可以通过分解成为 BCNF。将 STJ 分解成 ST（S，T）与 TJ（T，J），它们都是 BCNF。
- 一个模式中的关系模式如果属于BCNF，那么在函数依赖的范畴内，它已实现了彻底的分离，已消除了插入异常和删除异常。3NF的“不彻底”表现在可能存在主属性对码的部分依赖和传递依赖。

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

6.2.7 多值依赖

- 在函数依赖的范畴内，BCNF的关系模式已经满足设计的要求，但仍存在问题。
- 例1 学校中某一门课程由多个教员讲授，他们使用相同的参考书。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。可以用下面的表格来表示教员 T，课程 C 和参考书 B 之间的非规范化的关系及规范化的二维表格。

6.2.7 多值依赖

课 程 C	教 员 T	参 考 书 B
物理	李 勇	普通物理学
	王 军	光学原理
		物理习题集
数学	李 勇	数学分析
	王 军	微分方程
		高等代数
计算数学	张 平	数学分析
	周 峰	
.	.	.
.	.	.
.	.	.

6.2.7 多值依赖

课程C	教员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
.	.	.
.	.	.
.	.	.

6.2.7 多值依赖

- 可以定义关系模式为TEACHING（C，T，B），码是（C，T，B），即全码。因而TEACHING∈BCNF,不存在插入和删除异常，但对数据的增删改很不方便，数据的冗余也十分明显。其原因是存在多值依赖。

6.2.7 多值依赖

- **定义6.9** 设R（U）是属性集U上的一个关系模式。X，Y，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。

6.2.7 多值依赖

- 定义说明

  在关系模式TEACHING中，对于一个（物理，光学原理）有一组值{李勇，王军}，这组值仅仅决定于课程 C 上的值（物理）。对于另一个（物理，普通物理学）它对应的一组值仍是{李勇，王军}，尽管这时参考书 B 的值已经改变了。因此 T 多值依赖于 C，即C→→T。

6.2.7 多值依赖

- 多值依赖的等价定义

在R（U）的任一关系 r 中，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∈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值所得到的两个新元组必在 r 中），则 Y 多值依赖于 X，记作X →→ Y。这里，X，Y是 U 的子集，Z=U-X-Y。

若X →→ Y，而 Z 为空，则为平凡的多值依赖。

6.2.7 多值依赖

- 例2 关系模式WSC（W，S，C）中，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。则可以得到下表的关系模式：

6.2.7 多值依赖

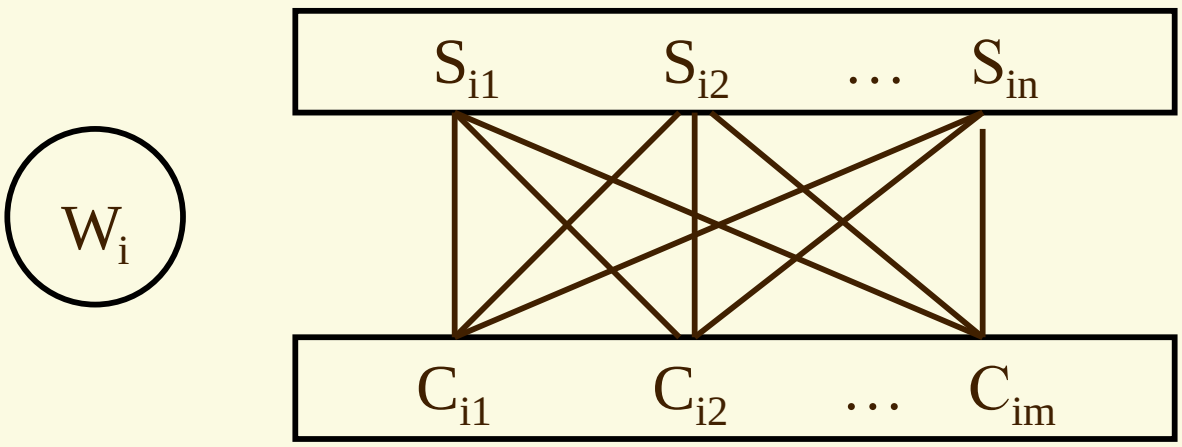
W	S	C
W1	S1	C1
W1	S1	C2
W1	S1	C3
W1	S2	C1
W1	S2	C2
W1	S2	C3
W2	S3	C4
W2	S3	C5
W2	S4	C4
W2	S4	C5

6.2.7 多值依赖

- 例2分析

对于W的每一个值W_i，S有一个完整的集合与之对应而不论C取何值。所以W→→S。

- 图示，为一个完全二分图，因而W→→S。由于C与S的完全对称性，必然有W→→C成立。



6.2.7 多值依赖

- 多值依赖具有以下性质：
  1. 对称性。若 X →→ Y，则 X →→ Z，其中Z＝U-X-Y。
  2. 传递性。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。因为当X→Y时，对 X 的每一个值 x ，Y 有一个确定的值 y 与之对应，所以X →→ Y。
  - 若X→→Y， X→→Z，则 X →→ YZ。
  - 若X→→Y， X→→Z，则 X →→ Y∩Z。
  - 若X→→Y， X→→Z，则 X →→ Y-Z， X →→ Z-Y 。

6.2.7 多值依赖

- 多值依赖与函数依赖的区别（了解）
  1. 多值依赖的有效性与属性集的范围有关。若X→→Y在U上成立则在W（XY ⊆ W ⊆ U）上一定成立；反之则不然。
  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。而多值依赖X →→ Y若在R（U）上成立，却不能断言对于任何Y'⊂ Y有X →→ Y'成立。

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

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不包含于X），X都含有码，则称R（U，F）∈4NF。
- 4NF就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，所允许的非平凡的多值依赖实际上是函数依赖。

6.2.8 4NF

- 如果一个关系模式是4NF，则必是BCNF。但BCNF不一定是4NF。上例中的关系模式WSC中，W →→ S，W →→ C，它们都是非平凡的多值依赖。而W不是码，（W，S，C）为码，因此WSC不是4NF。
- 属于BCNF但不属于4NF的关系模式，仍然具有不好的性质，数据的冗余度较大，应该继续规范化使其达到4NF。

6.2.8 4NF

- 可以用投影分解的方法消去非平凡且非函数依赖的多值依赖。上例中可以将WSC分解为WS（W，S），WC（W，C）。在WS中，W →→ S，但属于平凡的多值依赖，所以WS∈4NF，同理WC∈4NF。
- 函数依赖和多值依赖是两种最重要的数据依赖。还存在一种数据依赖，即连接依赖。函数依赖是多值依赖的一种特殊情况，而多值依赖又是连接依赖的特殊情况。如果消去了4NF中存在的连接依赖，则可以达到5NF的关系模式。

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

6.2.9 规范化小结

- 规范化的基本思想是逐步消除数据依赖中不合适的部分，是模式中的各关系模式达到某种程度的“分离”，即“一事一地”的模式设计原则。让一个关系描述一个概念、一个实体或者实体间的一种联系。若多余一个概念就把它“分离”出去。因此所谓的规范化实质上是概念的单一化。

6.2.9 规范化小结

- 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。