关系数据库数据理论新进展

郝忠孝 著

机械工业出版社

关系数据库数据理论新进展

郝 忠 孝 著

机械工业出版社

内 容 简 介

本书介绍了关系数据库数据理论近十几年来的新成果。主要内容有求候选关键字集和求全部候选关键字的几种方法、主属性的判定、函数依赖的推导、连接依赖的可满足性、初等范式、简单范式、BCNF 判定、Bernstein 算法的改进和有条件的综合算法等。

本书可供计算机专业的硕士、博士生及高级研究人员、高校教师及从事数据库研究工作的人员参考使用，也可作为数据库教学参考书。

图书在版编目(CIP)数据

关系数据库数据理论新进展/郝忠孝著．—北京：机械工业出版社，1998.4

ISBN 7-111-05999-9

Ⅰ.关… Ⅱ.郝… Ⅲ.关系数据库-数据库系统-理论

Ⅳ.TP311.13

中国版本图书馆 CIP 数据核字(97)第 22739 号

出版人：马九荣(北京市百万庄路 22 号；邮政编码 100037)

责任编辑：周保东　版式设计：周保东

封面设计：宋晓磊

北京市白帆印刷厂印刷 · 新华书店北京发行所发行

1998 年 1 月第 1 版第 1 次印刷

850mm×1168mm 1/32 · 6.25 印张 · 145 千字

0 001-1 200 册

定价：12.00 元

前　言

自从 1970 年 E. F. Codd 提出关系模型以来，关系数据库的理论和应用均得到了飞速发展。现在的关系数据库理论体系的源头始于 E. F. Codd 的论文："A Relational Model of Data for Large Shared Data Bankg"。接着又发表了一系列的关于关系数据库方面的有价值的论文。正是由于他的贡献，1980 年获得了图灵奖。他确立了关系数据库理论的框架。这一理论直到 80 年代初才算基本完成。其标志是 J. D. Ullman1980 年发表了专著："Principles of Database Systems"之后，C. J. Date 和 D. Maier 相继发表了有关专著，促进了关系数据库理论的发展。

关系数据库数据理论是关系数据库理论的核心。然而，有些问题并没有得到很好的解决，有些问题虽然解决了，但又很不彻底。因此，国内外的专家仍然继续研究这些问题。近十几年来，取得了丰硕的成果。作者有幸参加了部分研究工作，得到了国家和同行们的支持与帮助，对数据理论进行了探讨。本书正是在这一基础上撰写而成的。

我撰写本书的目的是想起到抛砖引玉的作用：使读者了解本书所涉及的问题现在已解决的程度，希望在以后的研究中取得更好的结果，推动数据理论的发展；为更好地解决关系数据库的应用和为其他类型数据库问题的解决提供一些条件；通过本书向读者提供一些解决数据库问题的有效方法。

全书共分五章。第一章为预备知识，是全书的基础；第二章是候选关键字和主属性问题，讨论了利用不同方法求候选关键字和主属性及其有关判定问题；第三章为数据依赖问题和最优覆盖问题，讨论了从关系中推导出函数依赖，讨论了有关模板依赖的几个问题，给出了最优覆盖的子类的求法；第四章为范式问题，讨论了

非经典范式的概念及其分解问题，相应的范式判定问题；第五章为规范化问题，讨论了规范化和候选关键字间的关系、Bernstein 算法的改进及有条件的综合算法等问题。

本书只纳入了近十年来关系数据库数据理论研究的部分成果，尚需要不断地充实和提高。我相信随着国内外同行的努力，关系数据库数据理论的研究将会取得更广泛、更完善的成果。在此基础上，将会有新的专著问世。

本书的内容多数是来自作者的国家自然科学基金资助项目，省自然科学基金重点资助项目的研究结果。在出版过程中得到了齐齐哈尔市科学技术委员会的帮助，在此一并表示感谢！对机械工业出版社为该书的出版做了大量工作的同志，表示深深的谢意！

该书由燕山大学刘国华、郭景峰先生主审，提出了不少宝贵意见，谢谢他所付出的劳动。

由于本书所涉及的内容绝大部分是新的，其理论和方法尚未来得及实践和验证，加之作者水平有限，书中难免有错误和不妥之处，敬请同行和读者批评指正，在此表示衷心感谢。

作者

1997 年 4 月

目　　录

第一章　预备知识

自 E. F. Codd 提出关系模型以来，关系数据库理论及构成该理论的重要组成部分的数据理论均得到了迅速发展。

关系模型、数据依赖、候选关键字约束、范式理论、规范化及模式分解等均属于关系数据库数据理论研究范畴。本章只介绍那些与本书相关的基础知识。

第一节　基本概念

一、关系模型和关系模式

关系模型是常见的三种数据模型(网状模型、层次模型和关系模型)中最重要的一种。关系模型是建立在严格的数学概念基础上的。在关系模型中，数据在用户观点下的逻辑结构是一张二维表。

按照关系模型建立的数据库是关系的有限集。一个关系就是一张二维表，表中的一行称为一个元组，表中一列的首由属性来表示。

域(Domain)是值的集合，是属性取值的范围。

给定一组域 $D_1, D_2, \cdots, D_n$ ($i \neq j$，但可以有 $D_i = D_j$)，其笛卡尔积为 $D_1 \times D_2 \times \cdots \times D_n = \{(d_1, d_2, \cdots, d_n) \mid d_i \in D_i, i = 1, 2, \cdots, n\}$，其中每一个元素称为一个 n 元组，或简称为元组，元素中的每一个值 d_i 称为一个分量。

若 $D_i (i = 1, 2, \cdots, n)$ 为有限集，其基数为 $m_i (i = 1, 2, \cdots, n)$，则笛卡尔积 $D_1 \times D_2 \times \cdots \times D_n$ 的基数为 $m = \prod_{i=1}^{n} m_i$。

现在可以给出关系、元组及属性等的严格数学定义。

定义 1-1　$D_1 \times D_2 \times \cdots \times D_n$ 的子集称为在域 $D_1, D_2, \cdots, D_n$ 上的关系，用 $R(D_1, D_2, \cdots, D_n)$ 表示，R 表示关系的名字，n 为关系的度或元。

每个元素是关系中的元组，用 t 表示。

表的每列对应一个域。由于域可以相同，为了区分，给每列起一个名字，称为属性。n 元关系必须有 n 个属性。

一个关系尽管是一个二维表，但这个二维表，必须具有如下限制：

(1)行、列的顺序无所谓。

(2)每一列中的分量是同一类数据，并出自同一个域。

(3)不同的列可出自同一个域，给予不同的属性名。

(4)任意两个元组不能全同。

(5)每一分量是不可分的数据项。

关系模型是自然界众多的应用环境的模型化的一种。它具有单一的数据结构——关系。在关系模型中，关系既表示实体集合，又表示实体间的联系。关系中的一个元组就表示一个实体，不同的元组表示不同的实体，同一类型的实体集用同一个关系的诸元组集合来表示。

由于按照关系模型构造的数据库是关系有限集，因此用户在数据库上的操作和查询也是在关系上定义和进行的。

上面已经说明，对一个二维表(关系)有诸多限制，这说明关系模型的关系必须满足相应的约束条件(完整性约束)。因此，说明关系模型的构成必须具备三要素：关系、操作和完整性约束。

关系模式是对关系中信息内容结构的描述。这里的信息结构包含属性、域、各属性间的关联关系及其他一些约束条件等。直观地说，关系模式是一个关系的属性名表(二维表的框架)。

关系是关系模式在某一时刻的实例。关系模型是型，是静态的；关系是值，是动态的，随着时间推移不断地变化。

关系模式的有限集构成数据库模式。

严格地说，一个关系模式应当是一个五元组，即 $R\langle U, D, DOM, \Sigma\rangle$。其中：$R$ 是关系名；U 是属性集；D 是属性集 U 中的属性所来自的域；DOM 是 U 到 D 的映射；Σ 是属性组的数据依赖的一组约束。

由于 D 和 DOM 对关系模式设计意义不大，因此，可以对关系模式做如下定义。

定义 1-2 一个关系模式是一个三元组 $R\langle U,\Sigma\rangle$，R 是关系名，U 是属性集，Σ 是 U 的数据依赖的一组约束。

如果只讨论在函数依赖约束下的关系模式，则记关系模式为 $R\langle U,F\rangle$。有时，习惯上也记为 $R(U)$。更甚者，讨论在其他依赖约束下有时也把关系模式记为 $R\langle U,F\rangle$，也有时记为 $R\langle U,D\rangle$等。

二、关系运算

关系模型的构成必须具备三要素。在关系模型下，对数据库的全部操作都被归结为对关系的运算。对关系的运算又都归结为集合运算，即以一个或多个关系作运算分量，经过运算，结果为一个新关系。

关系运算有两种不同的表达形式：关系代数、关系演算。最早在 1970 年，E. F. Codd 在文献[1]引入关系模型时就引入了关系代数。在 1972 年，E. F. Codd 在文献[2]中比较完整地讨论了关系代数的运算，并把它分成了两类：

(1)传统集合运算　并、交、差、笛卡尔积。

(2)专门集合运算　投影、选择、连接、自然连接、除。

所谓专门集合运算是为在数据库环境下进行相应的操作而设计的关系运算。

关系演算也是由 E. F. Codd 在文献[3]引入的另一种表达形式的关系运算。并给出了关系演算的两种类型：元组演算和域演算，同时还证明了关系代数和元组演算的等价性。元组演算和域演算的等价性在 1979 年被 J. D. Ullman 在文献[10]中所证明。

本书在后面的讨论中主要用到的是关系代数而不是关系演算。又由于关系代数和关系演算的等价性，则在本书中只介绍关系代数。

(一)传统集合运算

1. 并运算　两个关系 R 和 S 的并产生一个新关系，由属于 R 或属于 S 或同时属于 R 和 S 的所有元组构成，即

$$R \cup S = \{t \mid t \in R \vee t \in S\}$$

关系的并运算，要求参加运算的两个分量 R 和 S 必须是具有相同个数的属性，并且对应的列所代表的属性必须具有相同的域。

2. 差运算　两个关系 R 和 S 的差产生一个新关系，由属于 R 但不属于 S 的元组构成，即

$$R - S = \{t \mid t \in R \wedge t \notin S\}$$

关系的差运算，也要求参加运算的两个分量 R 和 S 必须是具有相同个数的属性，并且对应的列所代表的属性必须具有相同的域。

3. 交运算　两个关系 R 和 S 的交产生一个新关系，由既属于 R 又属于 S 的元组构成，即

$$R \cap S = \{t \mid t \in R \wedge t \in S\}$$

关系的交运算，也同样要求参加运算的两个分量 R 和 S 必须是具有相同个数的属性，并且对应的列代表的属性必须具有相同的域。

4. 笛卡尔积　一个 k_1 元关系 R 和一个 k_2 元关系 S 的笛卡尔积产生一个 (k_1+k_2) 元新关系。其中每一个元组的前 k_1 个分量是 R 中的某个元组，而后 k_2 个分量是 S 中的某个元组，即

$$R \times S = \{\langle a_1, \cdots, a_{k_1}, b_1, \cdots, b_{k_2}\rangle \mid \langle a_1, \cdots, a_{k_1}\rangle \in R \wedge \langle b_1, \cdots, b_{k_2}\rangle \in S\}$$

（二）专门集合运算

1. 投影运算　是从某个给定的关系中保留指定的属性子集而删去其余属性，产生一个属性子集。如果令 $R(X)$ 是给定的某个关系，X 是属性集，$Y \subseteq X$，则 $R(X)$ 在 Y 上的投影为

$$\pi_Y(R) = \{y \mid (\exists\, x \in R), y = x[Y]\}$$

式中，$x[Y]$ 为由元组 x 在 Y 所包含的那些属性上的分量所组成的子元组。

有时 $\pi_Y(R)$ 也记为 $R[Y]$。

值得注意的是，投影运算的结果属性子集中不能有重复元组，如有，则保留其中一个。

2. 选择运算　选择运算是从某个给定的关系中筛选出满足限定条件的元组子集，仍构成一个新关系，即

$$\sigma_F(R)=\{t \mid t\in R\wedge F\}$$

式中，F 为选择运算的限定条件的布尔表达式。

3. 连接运算　连接运算是从两个关系 R、S 的笛卡尔积中，选出属于 R 中某属性与 S 中某属性之间满足一定条件的那些元组，即

$$R\underset{A\theta B}{\geqslant} S=\sigma_{R\cdot A\theta S\cdot B}(R\times S)$$

式中，A、B 分别为 R 和 S 中的属性，θ 为算术比较运算符（$<$、$\leqslant$、$>$、$\geqslant$、$=$、$\neq$ 中任一运算符），$R\cdot A\theta S\cdot B$ 是满足连接的条件。

根据 θ 的不同，连接运算可以分为等值连接和不等值连接。

4. 自然连接运算　自然连接运算是从两个关系 R、S 的笛卡尔积中，选出同时属于 R、S 的同名属性值相等的那些元组。在产生的结果关系中同名属性也只出现一次，即

$$R(U_1)\geqslant S(U_2)=\{t \mid t[U_1]\in R\wedge t[U_2]\in S\}$$

其中，当 $U_1\cap U_2=\phi$ 时，即当 R 和 S 无公共属性时，$R\geqslant S$ 的结果与 $R\times S$ 相同。

5. 除运算　除运算是一个 k_1 元关系 R 除以一个 k_2 元关系 S，其中 $k_1>k_2$，$S\neq\phi$，R 中有 k_2 个属性与 S 的 k_2 个属性具有相同的域。从 R 中去掉 k_2 个与 S 相同的属性所余下的 k_1-k_2 个属性构成一个属性集。除运算的结果为出现在 S 中的所有元组都出现在 R 中的 k_1-k_2 余下的属性集，即

$$R\div S=\{t^{(k_1-k_2)} \mid \text{所有 } t^{(k_2)}\in S \text{ 均有 } t^{(k_1-k_2)}\cdot t^{(k_2)}\in R\}$$

以上讨论了九种运算，在这九种运算中自然连接运算是最重要的，这将在后面的讨论中可以看到，尤其是第四、五章的范式和规范化理论中更显示出其重要性。

这九种运算之间并不是互相独立的，某些运算可以通过其他运算来实现，并证明了并、差、笛卡尔积、选择、投影运算构成了关

系代数运算的最小完备集。

三、候选关键字

在讨论一个关系模式的数据依赖的约束之外，还有一类重要的约束是候选关键字约束。

正如上面的讨论一样，一个关系中任意两个元组不能相同，就限定了一个关系中每个元组的唯一性，因此，在任何一个关系中，都必定存在某个属性或属性集，它或它们的值唯一地标识一个元组，这个属性或属性集称为关系的候选关键字或超候选关键字。

定义 1-3 设 $R(U)$ 为关系模式，$U=A_1A_2\cdots A_n$，对于任一关系 $r\in R(U)$，属性集 $X\subseteq U$，当且仅当满足下列两条件时，X 为候选关键字。

(1)X 函数决定关系 r 的所有属性，即 $X\rightarrow A_i(i=1,2,\cdots,n)$。

(2)X 的任意真子集均不具有性质(1)，即若 $X'\subset X$，则 $X'\not\rightarrow A_i(i=1,2,\cdots,n)$。

若 X 具有性质(1)，但不具有性质(2)，则 X 称为超候选关键字。

可以看出，候选关键字除了具有唯一标识性之外，还具有无冗余性。

在给定的关系模式 $R(U)$ 中，对于关系 $r\in R(U)$，若不存在 $X\subseteq U$，$U=A_1A_2\cdots A_n$，使 $X\rightarrow A_i(i=1,2,\cdots,n)$，则 $U=A_1A_2\cdots A_n$ 为候选关键字，称为全候选关键字。这个关系 r 称为全候选关键字关系。

一个关系的候选关键字常常不唯一，常选择其中一个唯一的标识诸元组，被选中的候选关键字则称其为主候选关键字。

包含在任何一个候选关键字中的属性，称为该关系的主属性，反之，不属于任意一个候选关键字中的属性称为该关系的非主属性。

设 X 是关系 R 的一个属性集，X 不是关系 R 的候选关键字，但却是另一个关系 R_1 的候选关键字，则称 X 是 R 的一个外候选关键字。

外候选关键字是建立两个关系的中介。

四、几种数据依赖

一个数据依赖是关于关系数据库中关系的属性值之间的相关关系的一个命题，它规定了一个关系数据库的规范所满足的完整性约束条件。

满足数据依赖 d 所有关系的集合记作 SAT(d)。相应地，满足数据依赖 d 的一个关系 r 记作 $r\in$SAT(d)。

函数依赖是多种数据依赖中最常见、最重要的一类。

定义 1-4 设 $R(U)$ 是关系模式，U 是属性集，r 是 $R(U)$ 的任意一个关系，属性集 X、$Y\subseteq U$。对于任意两个元组 t_1、$t_2\in r$，有 $t_1[X]=t_2[X]$ 时，$t_1[Y]=t_2[Y]$。则称"Y 函数依赖于 X"或"X 函数决定 Y"，简记为 FD：$X\rightarrow Y$。

定义 1-5 如果 FD：$X\rightarrow Y$，但 $Y\not\subseteq X$，则称 FD：$X\rightarrow Y$ 是非平凡的函数依赖；如果 FD：$X\rightarrow Y$，但 $Y\subseteq X$，则称 FD：$X\rightarrow Y$ 是平凡的函数依赖。

定义 1-6 设 FD：$X\rightarrow Y$，如果对于任何的 $X'\subset X$，$X'\rightarrow Y$ 都不成立，则称 $X\rightarrow Y$ 是一个完全函数依赖，即 Y 函数依赖于整个 X，记为 $X\xrightarrow{f}Y$。

定义 1-7 设 FD：$X\rightarrow Y$，但不是完全函数依赖，则称 $X\rightarrow Y$ 是一个部分函数依赖，即 Y 函数依赖于 X 的某个真子集，记为 $X\xrightarrow{p}Y$。

定义 1-8 设 $R(U)$ 是关系模式，X、Y、$Z\subseteq U$。如果 $X\rightarrow Y$、$Y\rightarrow Z$ 且 $XY\cap Z=\phi$，但 $Y\not\rightarrow X$，则称属性集 Z 传递函数依赖于 X，记为 $X\xrightarrow{t}Z$。

以上几种数据依赖的定义均为函数依赖的范畴。由于函数依赖存在唯一性的特征，所以它表达了属性间的一对一的联系。但是，函数依赖很难表达出属性间一对多的联系，这就导致了多值依赖概念的提出。

为了给出多值依赖的定义，首先介绍像集的概念。

设 $R(U)$ 是一关系模式，关系 $r\in R(U)$，X、$Y\subseteq U$。显然，对于

元组 $t\in r$，t 中的一每一个 X 值 x，都存在着某个 Y 值的集合与之相关联，则称 Y 值的集合为 r 中 x 的 Y 像集，记作 $Y_r(x)$，即

$$Y_r(x)=\{t[Y]\mid t\in r\wedge t[X]=x\}$$

显然，当 r 满足 FD：$X\rightarrow Y$ 时，对 Y 中的任一 X 值 x，恒有 $|Y_r(x)|=1$，但实际上，$Y_r(x)$ 的基数因 x 而异。

定义 1-9 设 $R(U)$ 是一关系模式，X、$Y\subseteq U$，$Z=U-XY$，对于关系 $r\in R(U)$ 中的任意元组的每个 XZ 值 xz，都有 $Y_r(xz)=Y_r(x)$，即对于每一个给定的 XZ 值，其 Y 像集的值都仅仅依赖于该 XZ 值的 X 分量而与 Z 分量毫无关系。这时，称在 $R(U)$ 上 X 多值决定 Y，或称 Y 多值依赖于 X。记为 MVD：$X\rightarrow\rightarrow Y$。

显然，MVD：$X\rightarrow\rightarrow Y$ 与 MVD：$X\rightarrow\rightarrow Z$ 是对称的。

和 FD 一样，MVD 也存在某些平凡满足的问题，包含两种情况：

(1)当 $Y\subseteq X\subseteq\bigcup$ 时，$X\rightarrow\rightarrow Y$ 对于任何一个关系 $r(U)$ 是恒成立的。

(2)当 $U=XY$ 时，$X\rightarrow\rightarrow Y$ 对于任何 $r(U)$ 也是恒成立的。

把这两类 MVD 称为平凡多值依赖。

随着数据库理论及应用研究的进展，在讨论关系模式分解时，只考虑函数依赖和多值依赖是不够的，于是，便引进了连接依赖。

定义 1-10 设 $\rho=\{R_1,R_2,\cdots,R_m\}$ 是关系模式 $R(U)$ 的一个分解，一个关系 $r\in R(U)$，如果 $r=\pi_{R_1}(r)\bowtie\pi_{R_2}(r)\bowtie\cdots\bowtie\pi_{R_m}(r)$，即 r 可以无损连接地分解到 $R_1,R_2,\cdots,R_m$ 上，则称关系 r 满足连接依赖(简记为 JD)$\bowtie[R_1,R_2,\cdots,R_m]$。

五、数据依赖的逻辑蕴涵及推理规则

一般来说，数据依赖的关系不是孤立的，存在逻辑蕴涵问题。

定义 1-11 设 $R(U,\Sigma)$ 是一个关系模式，D 是 Σ 中的一个数据依赖集，d 为不同于 D 的数据依赖，如果每一个满足 D 中的全部数据依赖的关系也都满足 d，即 $SAT(D)\subseteq SAT(d)$，则称 D 逻辑蕴涵 d，记为 $D\models d$。

由于存在逻辑蕴涵，便出现了一些需要解决的问题：闭包问

题、成员籍问题、最简等价表示问题。

定义 1-12 设 $R\langle U,\Sigma\rangle$ 是关系模式，Σ 是数据依赖集，由 Σ 所逻辑蕴涵的所有的数据依赖称为 Σ 的闭包，记为 Σ^+。

定义 1-13 设 $R\langle U,\Sigma\rangle$ 是关系模式，Σ 是数据依赖集，要判定一个数据依赖 σ 是否属于 Σ^+，称为成员籍问题。

关于最简等价表示问题将在下一节中介绍。

为了解决这些问题和规范化分解和其他一些理论问题，需要研究有效的推理规则。W. W. Armstrong 于 1974 年在文献[7]中给出了在 FD 环境下的推理规则，称为 Armstrong 公理。

设 $R\langle U,F\rangle$ 是一个关系数据库的泛模式，F 是 U 上的 FD 集，$X,Y,Z,W\subseteq U$，则下述 Armstrong 推导规则可用于 FD 之间蕴涵关系的推导。

FD_1（自反规则）：对任何的 $Y\subseteq X\subseteq U$，$\vdash X\rightarrow Y$

FD_2（增广规则）：设 $Z\subseteq W$，则 $X\rightarrow Y\vdash XW\rightarrow YZ$

FD_3（传递规则）：$\{X\rightarrow Y,Y\rightarrow Z\}\vdash X\rightarrow Z$

FD_4（并规则）：$\{X\rightarrow Y,X\rightarrow Z\}\vdash X\rightarrow YZ$

FD_5（投影规则）：设 $Z\subseteq Y$，则 $X\rightarrow Y\vdash X\rightarrow Z$

FD_6（伪传递规则）：$\{X\rightarrow Y,WY\rightarrow Z\}\vdash WX\rightarrow Z$

规则中的符号 $\vdash$ 读为“可导出”。“$\vdash X\rightarrow Y$”表示 $X\rightarrow Y$ 是平凡可满足，因而 Armstrong 推理规则又称为 Armstrong 公理。

已经证明，Armstrong 公理是有效的和完备的。利用 Armstrong 公理可以进行某些理论推导，计算属性集闭包、成员籍判定算法等。除此之外，在覆盖与等价理论和数据库模式设计的综合算法等均必须用 Armstrong 公理。

有一个 Armstrong 公理的直接且很重要的结果是如下引理。

引理 1-1 设 $Y=A_1A_2\cdots A_n$，则 $X\rightarrow Y$ 成立当且仅当对于每一个 $A_i(i=1,2,\cdots,n)$，$X\rightarrow A_i$ 成立。

这一引理对于有关的理论的推理具有重要意义，将使许多问题的讨论极大地简化。甚至在不考虑这一引理结果的时候，常常会引起一些不应有的失误。这将在本书讨论 Bernstein 算法时看到这

一点。

现在给出一个很重要的概念，它不仅是证明 Armstrong 公理完备性的重要概念，而且在关系数据库数据理论中也起到很大作用，将在后面的讨论中看到。

定义 1-14 设 F 是在属性集 U 上的一个 FD 集，$X\subseteq U$，则 X 关于 FD 集 F 的属性集闭包 X_F 是所有这样的属性 A 的集合，只要 $X\rightarrow A$ 能从 F 借助于 Armstrong 公理导出。即

$X_F=\{A|X\rightarrow A$ 可由 Armstrong 公理从 F 导出$\}$

引理 1-2 $X\rightarrow Y$ 可由 Armstrong 公理从 F 导出的充要条件是 $Y\subseteq X^+$。

证明：充分性

设 $Y=A_1A_2\cdots A_n$，$Y\subseteq X_F^+$，根据定义 1-14，对于每个 $i(1\leqslant i\leqslant n)$，$F\vdash X\rightarrow A_i$，$A_i\in X^+$。

必要性

设 $F\vdash X\rightarrow Y$，且 $Y=A_1A_2\cdots A_n$，按投影规则知 $X\rightarrow A_i$ 成立。于是，$Y\subseteq X_F^+$。证毕。

需要指出的是这一引理为提供属性集闭包的求法奠定了基础。关于求解算法将在后面给出。

第二节　覆盖与等价

一、覆盖与等价

前面曾经指出，某些数据依赖可能被其他的数据依赖所逻辑蕴涵。因此，在讨论一个关系数据库中的数据依赖集时，常常涉及怎样尽可能且简捷地把它们表示出来的问题。

定义 1-15 设 $R\langle U,\Sigma\rangle$ 是一关系模式，U 为属性集，Σ 为数据依赖集。FD 集 F 和 G 均包含在 Σ 中，如果 $F^+=G^+$，则称 F 和 G 是等价的，记为 $F\equiv G$，如果 $F\equiv G$，则称 G 是 F 的一个覆盖。

两个等价的 FD 集在表示能力上是完全相同的。

根据引理 1-1，如果讨论的 FD 集中的每一个 FD 都是右部单