第六章

定义

关系：描述实体 属性 实体间的联系
形式上是一张二维表,是所有属性的笛卡尔积的子集
关系模式：用来定义关系
关系数据库：基于关系模式的数据库,用关系来描述现实世界
关系数据库的模式：定义这组关系的关系模式的全体

关系模式由5部分组成：R(U,D,DOM,F)

类型	解释
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	属性向域的映像集合
F	属性间数据的依赖关系集合

数据依赖

1.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
2.是现实世界属性间相互联系的抽象
3.是数据内在的性质
4.是语义的体现

数据依赖的类型

1.函数依赖(FD)
2.多值依赖(MVD)
3.其他

判断一个模式的好坏

会不会发生插入异常,删除异常,更新异常,数据冗余大
原因：存在数据依赖
解决办法：通过分解关系模式来消除其中不合适的数据依赖

函数依赖

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)上任何一个可能的关系r,r中不存在两个元组在X上相等而在Y上不等,则称X函数确定Y,记做X->Y.
X称为函数依赖的决定属性集
Y = f(X)

平凡函数依赖和非平凡函数依赖

如果 $X->Y$,但 $Y \nsubseteq X$ 则称 $X->Y$ 是非平凡的函数依赖
如果 $X->Y$,但 $Y \subseteq X$ 则称 $X->Y$ 是平凡的函数依赖

完全函数依赖和部分函数依赖

在关系模式R(U)中,如果 $X \rightarrow Y$,并且 X 的任何一个真子集 $X^{'}$ 都有 $X^{'} \nrightarrow Y$,则称Y完全函数依赖于X,记作 $X^{f} \rightarrow Y$,如果Y不完全函数依赖于 X ,则称 Y 部分函数依赖于 X,记作 $X^{P} \rightarrow Y$.
例如 : $Sno \nrightarrow Grade, Cno \nrightarrow Grade, \therefore (Sno, Cno)^{f} \rightarrow Grade$

传递函数依赖

在关系模式R(U)中,如果 $X \rightarrow Y, Y \rightarrow Z,$ 且$Y \nsubseteq X, Z \nsubseteq Y, Y \nrightarrow X$,则称 Z 传递函数依赖于 X.

码

设K为关系R<U,F>中属性或属性组合.若 $K^{f} \rightarrow U$,则称K为R的一个候选码.若关系模式R中有多个候选码,则选定其中一个作为主码.

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).
$1NF \in 2NF \in 3NF \in BCNF \in 4NF \in 5NF$

1NF

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为1NF
1NF是最起码的要求,不满足1NF的数据库模式不能成为关系数据库

2NF

若关系模式 $R \in 1NF$,并且每个非主属性都完全函数依赖于R的码,则 $R \in 2NF$

3NF

关系模式R<U,F>中如果不存在这样的码X,属性组Y以及非主属性Z($Z \nsubseteq Y$) 使得 $X \to Y, Y \nrightarrow X, Y \to Z$成立,则称 $R < U, F > \in 3NF$
在3NF中,每一个非主属性既不部分函数依赖于候选码也不传递函数依赖于候选码.

BCNF

设关系模式 $R < U, F > \in 1NF$,如果对于R的每个函数依赖 $X \to Y$,若 $Y \notin X$,则X必含有候选码,那么 $R \in BCNF$
特点:
1.每个决定属性集(因素)都包含(候选)码.
2.R中的所有属性(主,非主属性)都完全函数依赖于候选码.
3.所有主属性都完全函数依赖于每个不包含它的候选码.
4.没有任何属性完全函数依赖于非码的任何一组属性.
如果 $R \in 3NF$,且 R 只有一个候选码,则 R 必属于 BCNF

4NF

BCNF的不利 : 存在多值依赖
多值依赖 : 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，则称多值依赖
$X \to\to Y$ 成立当且仅当对R的任一关系 r， r 在（X， Z）上的每个值对应一组Y的值，这组值仅仅决定于X而与Z无关

平凡多值依赖和非平凡多值依赖

若 $X \to\to Y$,而 $Z = \emptyset$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
1.对称性 若 $X \to\to Y$,则 $X \to\to Z$,其中 Z＝U－X－Y
2.传递性 若 $X \to\to Y, Y \to\to Z,$ 则$X \to\to Z - Y$
3.函数依赖是多值依赖的特殊情况
若 $X \to Y,$ 则 $X \to\to Y.$
若 $X \to\to Y, X \to\to Z,$ 则 $X \to\to Y \bigcup Z.$
若 $X \to\to Y, X \to\to Z,$ 则 $X \to\to Y \bigcap Z.$
若 $X \to\to Y, X \to\to Z,$ 则 $X \to\to Y - Z, X \to\to Z - Y.$

多值依赖和函数依赖的区别

１．有效性
多值依赖的有效性与属性集的范围有关

多值依赖的定义不仅涉及属性组 X 和 Y，而且涉及 U 中其余属性 Z
2.
若函数以来在 R(U)上成立，则对任何 $Y^{'} \subset Y$ 均有 $X \to Y^{'}$
而多值依赖若有 R(U)上成立，不能断言对于任何 $Y^{'} \subset Y$有 X $\to\to Y^{'}$

4NF

关系模式$R < U, F > \in 1NF$,如果对于 R 的每个非平凡多值依赖$X \to\to Y (Y \not\subseteq X)$,X都含有候选码，则$R \in 4NF$
不允许有非平凡且非函数依赖的多值依赖
允许的是函数依赖(是非平凡多值依赖)

规范化

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集合，这种过程就叫关系模式的规范化
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

规范化的基本思想

消除不合适的数据依赖
使各关系模式达到某种程度的“分离”
所谓规范化实质上是概念的单一化
采用“一事一地”的模式设计原则

例题

引例: R (X , Y , Z)是第几范式?
$F = Y \to Z, XZ \to Y$
R 的侯选关键字为 XY 和 XZ , R 中所有属性都为主属性,不存在非主属性对侯选关键字的传递依赖,但存在主属性 Z 对关键字 XY 的部分依赖(也可看成传递依赖($XY \to Y, Y \to Z$) 或者说决定因素 Y 被关键字包含或决定因素不是关键字，所以仅为3NF
如何判定关系的范式?
-- 找出所有函数依赖,确定码

如何判定关系的码?

数据依赖的公理系统

逻辑蕴涵

关系模式 R<U,F>, F 是其函数依赖，X,Y是其属性子集，如果从F的函数依赖能够推出 $X \to Y$,则称 F 逻辑蕴涵 $X \to Y$记作$F \vdash X \to Y$
被 F 所逻辑蕴涵的函数依赖的全体所构成的集合称作 F 的闭包，记作$F^+ = \{X \to Y | F \vdash X \to Y\}$

Armstrong 公理系统

设 U 为属性集总体， F 是一组 U 上的函数依赖， X， Y， Z 是属性集，对R<U,F>有
自反律：若$Y \subseteq X$, 则$X \to Y$
增广律：若$X \subseteq Y$, 则$XZ \to YZ$
传递律：若$X \to Y, Y \to Z$, 则$X \to Z$
合并律：若$X \to Y, Y \to Z$, 则$X \to YZ$
分解律：若$X \to Y, Z \subseteq Y$, 则$X \to Y, X \to Z$
伪传递律：若$X \to Y, WY \to Z$, 则$WX \to Z$
有效性：用Armstrong公理从 F 中导出的函数依赖一定存在F^+
完备性：F^+中每一个函数依赖，必定可由 F 根据Armstrong公理推出来

属性集的闭包

设 F 为属性集 U 上的一组函数依赖，$X \subseteq U$,设$X_F^+ = \{A | X \to A 能根据 Armstrong 导出\}$，则称$X_F^+$为属性集 X 关于函数依赖集的闭包
只要该属性集的闭包包含关系的所有属性,且该属性集的每个真子集不包含关系的所有属性。那么该属性集就为候选码.

函数依赖集的等价性

1．函数依赖集 F， G，若$F^+ = G^+$,则称 F 与 G 等价
2．$F^+ = G^+ \leftrightarrow F \subseteq G^+, G \subseteq F^+$

最小覆盖

1．单属性化：F 中任一函数依赖$X \to A$,A必是单属性
逐个检查 F 中各函数依赖 $FD_i: X \to Y$,若$Y = A_1, A_2 \ldots A_k, k \geq 2$, 则用诸$X \to A_i$代替$Y$
2．无冗余化：F 中不存在依赖$X \to A$，使得 F 与$F - \{X \to A\}$等价.
逐个检查F中各函数依赖 $X \to A$, 令$G = F - \{X \to A\}$, 若$A \in X_G^+$,则从F中去掉该函数依赖
3．既约化：F 中不存在依赖$X \to A$，在 X 中有真子集 Z，使得 F 与$F - \{X \to A\} \bigcup \{Z \to A\}$
逐个检查F中各函数依赖$X \to A$,设$X = B_1 \ldots B_m$,逐个考查B_i,若$A \in (X - B_i)_F^+$,则以$(X - B_i)$取代X

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并不是唯一的，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.

等价定义

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

关系模式$R<U,F>$的一个分解：$\rho=\{R_1<U_1,F_1>,R_2<U_2,F_2>,...,R_n<U_n,F_n>\}$,其中$U=U_1\bigcup U_2\bigcup\cdots\bigcup U_n$, 且不存在$U_i\subseteq U_j$, F_i为 F 在U_i上的投影.

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

关系模式$R<U,F>$的一个分解：$\rho=\{R_1<U_1,F_1>,R_2<U_2,F_2>,...,R_n<U_n,F_n>\}$，若 R 与$R_1,R_2,...,R_n$自然连接的结果相等，则称关系模式 R 的这个分解$\rho$具有无损连接性.

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

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$U=U_1\bigcup U_2\bigcup\cdots\bigcup U_n$,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

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。

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