

第二章 关系数据库

本章系统地讲解了关系数据库的重要概念,并着重对关系模型进行了阐述。

关系模型包括关系数据结构、关系操作集合以及关系完整性约束三个组成部分。本章分别对这三个部分的内容进行了详细的分析与论述。

一、基本知识点

关系模型和关系数据库是《数据库系统概论》一书的重点,在全书中占有较大的篇幅(包括第二、三、四、五章)。因此,掌握本章的关键内容是学习后续各章的基础。

需要了解的:关系数据库理论产生和发展的过程,关系数据库产品的发展及沿革;关系演算的概念;域关系演算语言不包括在本科教学大纲内。

需要牢固掌握的:关系模型的三个组成部分及各部分所包括的主要内容;牢固关系数据结构及其形式化定义;关系的三类完整性约束的概念。

需要举一反三的:关系代数(包括抽象的语言及具体的语言);关系代数中的各种运算(包括并、交、差、选择、投影、连接、除及广义笛卡儿积等)、元组关系演算语言 ALPHA 及域关系演算语言 QBE 等,能够使用这些语言完成各种数据操纵。

难点:本章的难点在于关系代数。由于关系代数较为抽象,因此在学习的过程中一定要结合具体的实例进行学习。同时,要注意把握由具体语言到抽象语言的原则,即通过对具体语言如 ALPHA 和 QBE 的学习过渡到对抽象的关系演算的把握。

二、习题解答和解析

1. 试述关系模型的三个组成部分。

答

关系模型由关系数据结构、关系操作集合和关系完整性约束三部分组成。

2 试述关系数据语言的特点和分类。

答

关系数据语言可以分为三类:

关系代数语言 例如 ISBL

关系数据语言 关系演算语言 元组关系演算语言 例如 ALPHA, QUEL
域关系演算语言 例如 QBE

具有关系代数和关系演算双重特点的语言 例如 SQL

这些关系数据语言的共同特点是:具有完备的表达能力;是非过程化的集合操作语言;功能强;能够嵌入高级语言中使用。

3 定义并理解下列术语,说明它们之间的联系与区别:

(1) 域,笛卡儿积,关系,元组,属性

答

域:域是一组具有相同数据类型的值的集合。

笛卡儿积:给定一组域 D_1, D_2, \dots, D_n , 这些域中可以有相同的。这组域的笛卡儿积为

$$D_1 \times D_2 \times \dots \times D_n = \{ (d_1, d_2, \dots, d_n) \mid d_i \in D_i, i = 1, 2, \dots, n \}$$

其中每一个元素 (d_1, d_2, \dots, d_n) 叫做一个 n 元组 (n -tuple) 或简称元组 (Tuple)。元素中的每一个值 d_i 叫做一个分量 (Component)。

关系:在域 D_1, D_2, \dots, D_n 上笛卡儿积 $D_1 \times D_2 \times \dots \times D_n$ 的子集称为关系,表示为

$$R(D_1, D_2, \dots, D_n)$$

元组:关系中的每个元素是关系中的元组。

属性:关系也是一个二维表,表的每行对应一个元组,表的每列对应一个域。由于域可以相同,为了加以区分,必须对每列起一个名字,称为属性 (Attribute)。

(2) 主码,候选码,外部码

答

候选码:若关系中的某一属性组的值能惟一地标识一个元组,则称该属性组为候选码 (Candidate key)。

主码:若一个关系有多个候选码,则选定其中一个为主码 (Primary key)。

外部码:设 F 是基本关系 R 的一个或一组属性,但不是关系 R 的码,如果 F 与基本关系 S 的主码 K_s 相对应,则称 F 是基本关系 R 的外部码 (Foreign key),简称外码。

基本关系 R 称为参照关系 (Referencing relation), 基本关系 S 称为被参照关系 (Referenced relation) 或目标关系 (Target relation)。关系 R 和 S 可以是相同的关系。

(3) 关系模式,关系,关系数据库

答

关系模式:关系的描述称为关系模式 (Relation Schema)。它可以形式化地表示为

$$R(U, D, \text{dom}, F)$$

其中 R 为关系名, U 为组成该关系的属性名集合, D 为属性组 U 中属性所来自的域, dom 为属性向域的映像集合, F 为属性间数据的依赖关系集合。

关系:在域 D_1, D_2, \dots, D_n 上笛卡儿积 $D_1 \times D_2 \times \dots \times D_n$ 的子集称为关系, 表示为

$$R(D_1, D_2, \dots, D_n)$$

关系是关系模式在某一时刻的状态或内容。关系模式是静态的、稳定的;而关系是动态的、随时间不断变化的,因为关系操作在不断更新着数据库中的数据。

关系数据库:关系数据库也有型和值之分。关系数据库的型也称为关系数据库模式,是对关系数据库的描述,它包括若干域的定义以及在这些域上定义的若干关系模式。关系数据库的值是这些关系模式在某一时刻对应的关系的集合,通常就称为关系数据库。

4 试述关系模型的完整性规则。在参照完整性中,为什么外部码属性的值也可以为空?什么情况下才可以为空?

答

关系模型的完整性规则是对关系的某种约束条件。关系模型中可以有三类完整性约束:实体完整性、参照完整性和用户定义的完整性。

其中实体完整性和参照完整性是关系模型必须满足的完整性约束条件,被称做是关系的两个不变性,应该由关系系统自动支持。

(1) 实体完整性规则:若属性 A 是基本关系 R 的主属性,则属性 A 不能取空值。

(2) 参照完整性规则:若属性(或属性组) F 是基本关系 R 的外码,它与基本关系 S 的主码 K_s 相对应(基本关系 R 和 S 不一定是不同的关系),则对于 R 中每个元组在 F 上的值必须为:

- 1) 或者取空值(F 的每个属性值均为空值);
- 2) 或者等于 S 中某个元组的主码值。

(3) 用户定义的完整性是针对某一具体关系数据库的约束条件。它反映某一具体应用所涉及的数据必须满足的语义要求。

在参照完整性中,外部码属性的值可以为空,它表示该属性的值尚未确定,但前提条件是该外部码属性不是其所在关系的主属性。

例如,在下面的“学生”表中,“专业号”是一个外部码,不是学生表的主属性,可以为空,其语义是,该学生的专业尚未确定。

学生(学号, 姓名, 性别, 专业号, 年龄)

专业(专业号,专业名)

而在下面的“选修”表中的“课程号”虽然也是一个外部码属性,但它又是“课程”表的主属性,所以不能为空,因为关系模型必须满足实体完整性。

课程(课程号,课程名,学分)

选修(学号,课程号,成绩)

5 设有一个 SPJ 数据库,包括 S、P、J、SPJ 四个关系模式:

S(SNO, SNAME, STATUS, CITY);

P(PNO, PNAME, COLOR, WEIGHT);

J(JNO, JNAME, CITY);

SPJ(SNO, PNO, JNO, QTY);

供应商表 S 由供应商代码 (SNO)、供应商姓名 (SNAME)、供应商状态 (STATUS)、供应商所在城市 (CITY) 组成;

零件表 P 由零件代码 (PNO)、零件名 (PNAME)、颜色 (COLOR)、重量 (WEIGHT) 组成;

工程项目表 J 由工程项目代码 (JNO)、工程项目名 (JNAME)、工程项目所在城市 (CITY) 组成;

供应情况表 SPJ 由供应商代码 (SNO)、零件代码 (PNO)、工程项目代码 (JNO)、供应数量 (QTY) 组成,表示某供应商供应某种零件给某工程项目的数量为 QTY。

今有若干数据如下:

S 表			
SNO	SNAME	STATUS	CITY
S1	精 益	20	天津
S2	盛 锡	10	北京
S3	东方红	30	北京
S4	丰泰盛	20	天津
S5	为 民	30	上海

P 表			
PNO	PNAME	COLOR	WEIGHT
P1	螺 母	红	12
P2	螺 栓	绿	17
P3	螺丝刀	蓝	14
P4	螺丝刀	红	14
P5	凸 轮	蓝	40
P6	齿 轮	红	30

J 表

JN0	JNAME	CITY
J1	三 建	北京
J2	一 汽	长春
J3	弹簧厂	天津
J4	造船厂	天津
J5	机车厂	唐山
J6	无线电厂	常州
J7	半导体厂	南京

SPJ 表

SN0	PN0	JN0	QTY
S1	P1	J1	200
S1	P1	J3	100
S1	P1	J4	700
S1	P2	J2	100
S2	P3	J1	400
S2	P3	J2	200
S2	P3	J4	500
S2	P3	J5	400
S2	P5	J1	400
S2	P5	J2	100
S3	P1	J1	200
S3	P3	J1	200
S4	P5	J1	100
S4	P6	J3	300
S4	P6	J4	200
S5	P2	J4	100
S5	P3	J1	200
S5	P6	J2	200
S5	P6	J4	500

试分别用关系代数、ALPHA 语言、QBE 语言完成下列操作：

(1) 求供应工程 J1 零件的供应商号 SN0；

答

关系代数

SN0 (JN0 = J1 (SPJ))

ALPHA 语言

GET W (SPJ.SNO): SPJ.JNO = J1

QBE 语言

SPJ	SNO	PNO	JNO	QTY
	P. <u>S1</u>		J1	

(2) 求供应工程 J1 零件 P1 的供应商号 SNO;
答

关系代数

$SNO (\quad JNO = J1 \quad PNO = P1 (SPJ))$

ALPHA 语言

GET W (SPJ.SNO): SPJ.JNO = J1 SPJ.PNO = P1

QBE 语言

SPJ	SNO	PNO	JNO	QTY
	P. <u>S1</u>	P1	J1	

(3) 求供应工程 J1 红色零件的供应商号 SNO;
答

关系代数

$SNO (\quad SNO, PNO (\quad JNO = J1 (SPJ)) \quad PNO (\quad COLOR = 红 (P)))$

ALPHA 语言

RANGE P PX

GET W (SPJ.SNO): v PX (PX.PNO = SPJ.PNO SPJ.JNO = J1 PX .COLOR = 红)

QBE 语言

SPJ	SNO	PNO	JNO	QTY
	P. <u>S1</u>	<u>P1</u>	J1	

P	PNO	PNAME	COLOR	WEIGHT
	<u>P1</u>		红	

(4) 求没有使用天津供应商生产的红色零件的工程号 JNO;
答

关系代数

$JNO(J) - JNO(SNO(CITY = 天津(S)) \quad SNO, PNO, JNO(SPJ) \quad PNO(COLOR = 红(P)))$

解析

减法运算中被减的部分是使用了天津供应商生产的红色零件的所有工程号， $JNO(J)$ 是全部工程的工程号，两者相减就是没有使用天津供应商生产的红色零件的工程号，包括没有使用任何零件的工程号。

ALPHA 语言

```
RANGE SPJ      SPJX
      P      PX
      S      SX
GET W (J.JNO):  v SPJX (SPJX .JNO = J.JNO
                  v SX ( SX.SNO = SPJX .SNO  SX .CITY = 天津
                  v PX (PX .PNO = SPJX .PNO  PX .COLOR = 红 ))
```

解析

- 1) S、P、SPJ 表上各设了一个元组变量。
- 2) 解题思路是：要找的是满足给定条件的工程号 JNO，因此，对工程表 J 中的每一个 JNO 进行判断：看 SPJ 中是否存在这样的元组，其 JNO = J.JNO，并且所用的零件是红色的，该零件的供应商是天津的。

如果 SPJ 中不存在这样的元组，则该工程号 JNO 满足条件，放入结果集中。
如果 SPJ 中存在这样的元组，则该工程号 JNO 不满足条件，不放入结果集中。再对工程表 J 中的下一个 JNO 进行同样的判断。
直到所有 JNO 都检查完。

结果集中是所有没有使用天津供应商生产的红色零件的工程号，包括没有使用任何零件的工程号。

QBE 语言

当不考虑没有使用任何零件的工程时

S	SNO	SNAME	STATUS	CITY
	<u>S1</u>			天津

P	PNO	PNAME	COLOR	WEIGHT
	<u>P1</u>		红	

SPJ	SNO	PNO	JNO	QTY
	<u>S1</u>	<u>P1</u>	P.J1	

解析

本题是从 SPJ 表中输出满足条件的 JN0, 没有使用任何零件的工程项目的工程号是不会出现在 SPJ 中的, 所以本题的结果不包括没有使用任何零件的工程号。

考虑没有使用任何零件的工程

J	JN0	JNAME	CITY	
	P. <u>J1</u>			

S	SNO	SNAME	STATUS	CITY
	<u>S1</u>			天津

P	PNO	PNAME	COLOR	WEIGHT
	<u>P1</u>		红	

SPJ	SNO	PNO	JN0	QTY
	<u>S1</u>	<u>P1</u>	<u>J1</u>	

解析

本题是从 J 表中输出满足条件的 JN0, 没有使用任何零件的工程项目的工程号也满足条件, 所以本题的结果包括没有使用了使用任何零件的工程号。

(5) 求至少用了 S1 供应商所供应的全部零件的工程号 JN0。

答

关系代数

$$JN0, PNO (SPJ) \div PNO (SNO = S1 (SPJ))$$

解析

上面公式中除号前的部分是所有工程与该工程所用的零件, 除号后的部分是 S1 所供应的全部零件号。对于 SPJ 表中的某一个 JN0, 如果该工程使用的所有零件的集合包含了 S1 所供应的全部零件号, 则该 JN0 符合本题条件, 在除法运算的结果集中。可以看到, 使用关系代数的除法运算概念清晰, 语言表达也很简单。

ALPHA 语言 (类似于《概论》P67 例 14)

RANGE SPJ SPJX


```
SPJ SPJY
P      PX
GET W(J.JNO): " PX(v SPJX(SPJX.PNO = PX.PNO  SPJX.SNO = S1 )
                v SPJY(SPJY.JNO = J.JNO  SPJY.PNO = PX.PNO))
```

解析

1) SPJ 表上设了两个元组变量: SPJX, SPJY; P 表上设了一个元组变量: PX。

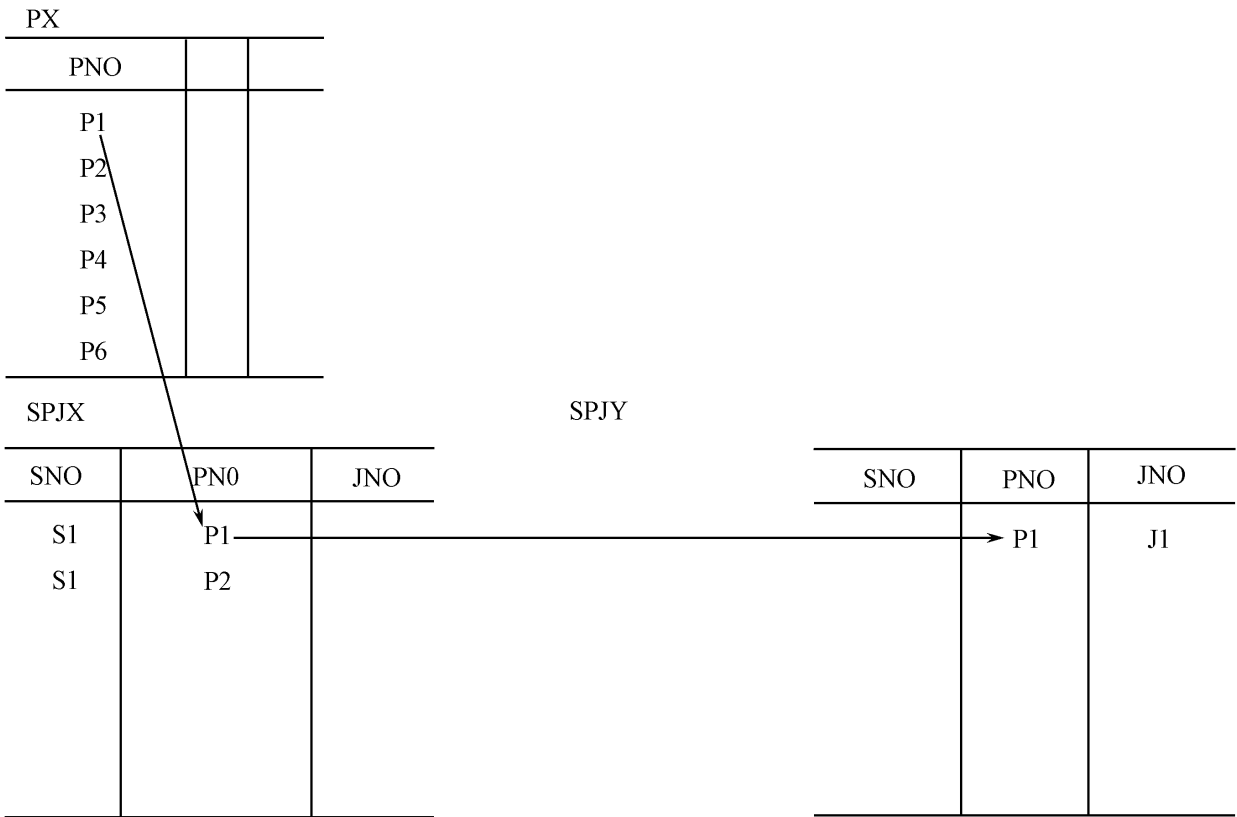
2) 解题思路是: 要找的是满足给定条件的工程号 JNO。因此, 对工程表 J 中的每一个 JNO(例如 J1), 进行以下一组操作:

对零件 PX 中的所有零件, 依次对每一个零件, 进行以下检查;

例如零件 P1, 检查 SPJX, 看 S1 是否供应了该零件, 如果供应了, 则再看这一个 JNO(例如 J1)是否使用了该零件;

如果对于 S1 所供应的每种零件, 这一个 JNO(例如 J1)都使用了, 则该 JNO 为(例如 J1)满足要求的工程项目。

3) 为了帮助理解, 读者可以画出所涉及三个表, 给出一些数据, 按照上面的解析步骤一步一步地分析, 就能掌握解题方法, 从而达到举一反三的要求。



QBE(不要求)。

6 试述等值连接与自然连接的区别和联系。

答

连接运算中有两种最为重要也最为常用的连接, 一种是等值连接 (Equijoin), 另一种是自然连接 (Natural join)。

为“ = ”的连接运算,称为等值连接。

它是从关系 R 与 S 的笛卡儿积中选取 A、B 属性值相等的那些元组。即等值连接为

$$R \underset{A=B}{\bowtie} S = \{ t_r t_s \mid t_r \in R, t_s \in S, t_r[A] = t_s[B] \}$$

自然连接(Natural join)是一种特殊的等值连接,它要求两个关系中进行比较的分量必须是相同的属性组,并且在结果中把重复的属性列去掉。即若 R 和 S 具有相同的属性组 B,则自然连接可记为

$$R \underset{B}{\bowtie} S = \{ t_r t_s \mid t_r \in R, t_s \in S, t_r[B] = t_s[B] \}$$

7. 关系代数的基本运算有哪些?如何用这些基本运算来表示其他的运算?
答

在八种关系代数运算中,并、差、笛卡儿积、投影和选择五种运算为基本的运算。其他三种运算,即交、连接和除,均可以用五种基本运算来表达。

$$\text{交运算: } R \cap S = R - (R - S)$$

$$\text{连接运算: } R \underset{A \cup B}{\bowtie} S = \underset{A \cup B}{\pi} (R \times S)$$

$$\text{除运算: } R(X, Y) \div S(Y, Z) = \underset{X}{\pi} (R) - \underset{X}{\pi} (\underset{X}{\pi} (R) \times \underset{Y}{\pi} (S) - R)$$

X、Y、Z 为属性组, R 中的 Y 和 S 中的 Y 可以有不同的属性名,但必须出自相同的域集。

三、小 结

关系数据库系统是目前使用得最为广泛的数据库系统,1970 年以后开发的数据库系统几乎都是基于关系的。可以说,关系模型的提出是数据库发展史上最为重要的成就。

本章对关系模型的核心概念进行了阐述,并重点讲述了关系数据结构、关系操作集合以及关系完整性约束的概念。对以上三个部分的掌握是学习后续章节的基础。

学习本章的关键在于理解关系模型的核心与本质,而不要仅仅局限于理解相关概念的字面意思。读者应该达到能够对所学知识灵活运用的程度。