

第二节 关系数据库模型



知识点



关系数据库的基本概念

关系代数

数据查询

关系数据库模型

基本概念

关系数据库的概念

关系、记录

- 一个表格就是一个关系。表中的每一行称为一个记录，也称为“元组”。

域

- 数据项的取值范围。

候选码

- 关系中唯一标识一个记录的属性或属性组。

主码

- 一个关系中候选码可能有多，选定一个作为主码。

学生基本信息表

姓名	学号	性别	学院	出生年月
李明	20140123	男	计算机学院	1992、7
王芳	20140125	女	外文学院	1990、8
...

关系

记录

候选码可由单个属性构成

主码

域：{1-12}

也可由多个属性组合

关系数据库模型

关系的性质

关系数据库的概念

表中的关键字属性必须有**唯一**的名字。

行的顺序**不是**影响因素。

任意两个记录**不能**完全相同。

表格**不能**嵌套。

每一列中分量是**同一类型**的数据。

学生表

姓名	学号	性别	学院	专业	出生年月	
					年	月
李明	20140123	男	计算机学院	软件工程	1991	6
李明	20140124	男	计算机学院	软件工程	1991	6
王芳	20140125	女	外文学院	英国文学	1990	7
陈红	20140125	男	数学学院	应用数学	1992	7
...		

可以只要主键不同

可以调换顺序

不能相同

来自同一个域

不允许嵌套

关系数据库模型

关系数据库的概念

关系的完整性

实体完整性

- 所有主码对应的属性值不能为空

学生表

姓名	学号	学院
李明	20140123	地理学院
王芳	20140125	计算机
...

不能为空

参照完整性

- 定义外码与主码之间的引用规则，即外码必须是另一个表的主码的有效值

课程表

课程号	课程名	学分
1	C#	3
2	GIS	4
...

主码

用户自定义完整性

- 针对某一具体关系数据库的约束条件

成绩表

学号	课程号	成绩
20140123	1	80
20140125	2	95
...

外码

0-100

关系数据库模型

传统的关系运算

关系代数是一种抽象的查询语言，用关系的运算来表达查询，是作为研究关系数据语言的数学工具。

关系代数

R

A	B	C
a ₁	b ₁	c ₁
a ₁	b ₂	c ₂
a ₂	b ₂	c ₁

S

A	B	C
a ₁	b ₂	c ₂
a ₁	b ₃	c ₂
a ₂	b ₂	c ₁

R ∪ S

A	B	C
a ₁	b ₁	c ₁
a ₁	b ₂	c ₂
a ₂	b ₂	c ₁
a ₁	b ₃	c ₂

R - S

A	B	C
a ₁	b ₁	c ₁

并 (Union)

由属于R或属于S的记录组成。

差 (Difference)

由属于R而不属于S的所有记录组成。

关系数据库模型

传统的关系运算

关系代数

列 : $m=3$
行 : $k1=3$

R

A	B	C
a_1	b_1	c_1
a_1	b_2	c_2
a_2	b_2	c_1

$R \cap S$

A	B	C
a_1	b_2	c_2
a_2	b_2	c_1

交 (Intersection)

由既属于R又属于S的记录组成

两个关系都有n个属性，相应的属性取自同一个域。

列 : $n=3$
行 : $k2=3$

S

A	B	C
a_1	b_2	c_2
a_1	b_3	c_2
a_2	b_2	c_1

$R \times S$

RA	RB	RC	SA	SB	SC
a_1	b_1	c_1	a_1	b_2	c_2
a_1	b_1	c_1	a_1	b_3	c_2
a_1	b_1	c_1	a_2	b_2	c_1
a_1	b_2	c_2	a_1	b_2	c_2
...			

笛卡尔积

列 : $(m+n)$ 列
行 : $k1 \times k2$ 个元组

$m+n=6$
 $k1 \times k2=9$

关系数据库模型

专门的关系运算

选择 (Selection)

选择又称为限制,在关系R中选择满足给定条件的诸元组构成新的关系：

$$\sigma_F(R) = \{ t \mid t \in R \wedge F(t) = \text{'真'} \}$$

F：选择条件，是一个逻辑表达式。

关系代数

学生表

姓名	学号	所在的系	性别
李明	20140123	测绘工程	男
王芳	20140125	GIS	女
陈强	20140126	GIS	男
张立	20140127	测绘工程	男

查询信息系（GIS系）全体学生

$\sigma_{\text{所在的系} = \text{'GIS'}}(\text{学生表})$

姓名	学号	所在的系	性别
王芳	20140125	GIS	女
陈强	20140126	GIS	男

- 选择是从行的角度进行的运算

关系数据库模型

专门的关系运算

投影 (Projection)

从关系中选择出若干属性列组成新的关系

$$\Pi_A(R) = \{ t[A] \mid t \in R, A : R \text{ 中的属性列} \}$$

关系代数

学生表

姓名	学号	所在的系	性别
李明	20140123	RS	男
王芳	20140125	GIS	女
王芳	20140126	GIS	男
赵金	20140127	RS	男

查询学生的姓名和所在系

姓名	所在的系
李明	RS
王芳	GIS
王芳	GIS
赵金	RS

投影操作主要是从列的角度进行运算

但投影之后不仅取消了原关系中的某些列
而且还可能取消某些元组 (避免重复行)

关系数据库模型

专门的关系运算

等值连接 (Join)

从关系R与关系S的广义笛卡尔积中选取R、S属性值相等的那些元组。

$$R \bowtie_{A=B} S = \{ \widehat{t_r t_s} \mid t_r \in R \wedge t_s \in S \wedge t_r[A] = t_s[B] \}$$

R

A	B	C
a ₁	b ₁	5
a ₁	b ₂	6
a ₂	b ₃	8
a ₂	b ₄	12

S

B	E
b ₁	3
b ₂	7
b ₃	10
b ₃	2
b ₅	6

R \bowtie S
R.B=S.B

A	R.B	C	S.B	E
a ₁	b ₁	5	b ₁	3
a ₁	b ₂	6	b ₂	7
a ₂	b ₃	8	b ₃	10
a ₂	b ₃	8	b ₃	2

一般的连接操作是从行的角度进行运算。
但还需要取消重复列。

重复列

关系数据库模型

数据库查找程序

数据查询

通常，查找一个数据库记录是一个连续的过程。计算机从头开始进行查询，直到找到需要的项目为止。

查找ID= 050792的地块信息

依次对比， 查询	地块ID	地块类型	调查人	调查时间
	≠ 010682	耕地	陈强	2012.10
	≠ 010683	建筑用地	李菲	2012.10
	≠ 010684	林地	陈强	2012.10

	050792	建筑用地	王强	2012.11
	050793	建筑用地	高磊	2013.10

查找成功

● 方法可行，如果记录较多，需要很长的时间才能找到信息，效率相当低。

关系数据库模型

二分法查找程序

数据查询

如果已经对数据进行了排序，数据查找效率就可得到提高。
二分法查找程序是计算机快速查找数据的典型例子。

算法步骤：

- 把资料组分为两部分
- 记录在前半部分还是在后半部分？
- 放弃不需要的一半；
- 重复上面的操作。

举例：查找数字11

4	10	11	25	42	59	64	71	73
①	②	③	④	⑤	⑥	⑦	⑧	⑨

↑ 11 < 42

4	10	11	25
①	②	③	④

↑ 11 > 10

11	25
③	④

↑ 11 = 11
查找成功程序结束

关系数据库模型

主要的算符

数据查询

- 数据库查询中所用到的算符包括：

Age >= 18 Sal = Sal + 1500

Name Lk 'g%' Age > 18 AND sex = '男'

标准逻辑算符：→

数学算符：→

字符处理运算符：→

布尔逻辑算符：→

=	<>	>	<	>=	<=	
+	-	*	/	7	8	9
CN	NC	LK	IN	4	5	6
AND	OR	NOT	0	1	2	3

关系数据库模型

结构化查询语言

为实现数据库查询而专门设计的一种标准化语言。应用广泛，可以进行独立的、分散的数据库连接。

数据查询



● SQL
(Structural Query
Language)

查询语言

广州市人口？

1600

广州市人均GDP？

4900

.....

全国各城市人口统计表

城市代码	城市名称	城市人口数
14536	北京	1300
14661	广州	1600
14663	上海	
...	...	

全国各城市
GDP总量

城市代码	城市名称	GDP总量
14536	北京	5600
14661	上海	6300
14663	广州	4900
...

关系数据库模型

SQL数据库查询的句法

SQL查询的基础句法是：

SELECT {数据字段}
FROM {表格}
WHERE <字段> Expression <值>

数据
查询

土地信息表

土地代码	面积	所有者	是否登记
14536	350	S.Engel	是
14661	399	G.Leeds	否
14663	399	S.Engel	是

例如：

SELECT 土地代码, 面积, 所有者
FROM 土地信息表
WHERE 面积 > 360

查询结果

土地代码	面积	所有者
14661	399	G.Leeds
14663	399	S.Engel

关系数据库模型

SQL连接查询示例

选择所有满足下列要求的土地：价值大于25000，其上有多于或等于5个房产：

SELECT 编号, 价值, 房产个数

FROM 土地表, 位置表

WHERE [价值 > 25000 AND 房产个数 >= 5)

查询条件

[AND (土地表.编号 = 位置表.编号)]

联接条件

土地表

编号	拥有者	价值	用途
34651	D.Smith	25000	Residential
34652	C.Lawrence	22500	Industrial
34653	D.Lewsley	32600	Agricultural
34654	S.Bishop	24300	Residential
34655	C.Lawrence	88500	Commercial

位置表

编号	地址	面积	房产个数
34651	George St.	1540	12
34652	Adam St.	2200	03
34653	Henry St.	9500	01
34654	Park Ave.	0950	03
34655	Station Rd.	1450	10

输出的
结果

编号	价值	房产个数
34651	25000	12
34655	88500	10



谢谢大家！

