DataBase

数据模型

概念模型

也称信息模型，它是按照用户的观点来对数据和信息建模，用于数据库设计。

逻辑模型和物理模型

- 逻辑模型
  - 主要包括网状模型，层次模型，关系模型，面向对象数据模型，对象关系数据模型，半结构化数据模型。
  - 按计算机系统的观点对数据建模，用于DBMS实现
- 物理模型

```
现实世界->认识抽象->概念模型->逻辑模型->物理模型
```

就如面向对象中对现实世界的抽象---对象，同一类型的对象的类型抽象出来就是类,通过类进行实例化产生对象。然而现实世界中并不存在"类"这种东西。通过建立模型来表达事物。实体的描述角度就叫做一个属性。实体型，同一类个体的共同属性定义下来就是一个型，描述的物体构成一个集合即是实体集。实体之间的联系通常是指不同实体集之间的联系。实体之间的联系有一对一，一对多，多对多等多种类型。

E-R方法

- E-R图来描述现实世界的概念模型
- E-R方法也称为E-R模型

如何描述一个模型

- 数据结构
  - 描述的内容有两类
    - 与对象的类型,内容性质有关
    - 与数据之间联系有关的对象
  - 对系统静态特性的描述
- 数据操作
  - 对系统动态特性的描述
- 完整性约束条件(integrity constraints)
  - 保证数据的正确,有效和兼容

关系模型

关系模型的数据结构就是关系，关系是一个集合。关系数据库系统采用关系模型作为数据的组织方式。

```
+ 数据结构
        + 关系：元组的集合
+ 数据操作
        + 关系代数
```

```
        + 关系演算
        + SQL
+ 数据完整性约束
        + 实体完整性约束
        + 参照完整性约束
        + 用户定义是完整性约束
```

在用户观点下关系模型中的数据的逻辑结构是一张二维表，它由行和列组成。

学生登记表：

学号	姓名	年龄	性别
2017003	Jason	19	M
2017004	David	20	M

```
关系：一个关系对应一张表。
元组：表中的一行即为一个元组
属性：表中的一列就是一个属性
码：也称为键，表中的某个属性组，可以唯一确定一个元组
域：是一组具有相同数据类型的值的集合。
分量：元组的一个属性值
关系模式：关系模式要求关系必须是规范化的，关系的每一个分量必须是一个不可分的数据项，不允许表中还有表。关系名（属性1，属性2....，属性n）
```

关系中的元组也是不可重复的。

- 优点
  - 建立在严格的数学概念的基础上
  - 概念单一
    - 实体和各类联系都用关系来表示
    - 对数据的检索结果也是关系
  - 关系模型的存取路径对用户透明
    - 具有更高的数据独立性，更好的安全保密性
    - 简化了程序员的工作和数据库开发建立的工作
- 缺点
  - 由于非过程化，查询效率不如格式化模型
  - 为提高性能，必须优化

术语对比

数据库结构语言SQL

SQL是一种结构化查询语言，是关系数据库的标准语言。是一个通用的功能极强的数据库语言。

SQL的特点

1. 综合统一

2. 高度非过程化
3. 面向结合的操作方式
4. 已同一种语法提供多种使用方式
5. 语言简介，易学易用

SQL的动词：

SQL功能	动词
数据查询	SELECT
数据定义	CREATE,DROP,ALTER
数据操纵	INSERT,UPDATE,DELETE
数据控制	CRANT,REVOKE

SQL的基本概念

插入数据

两种插入数据的方式

- 插入元组
- 插入子查询结果
  - 可以一次插入多个元组

插入元组

```
INSERT [INTO] <表名>[(<属性列1>[(<属性列2>)]] VALUES [<常量1>[(<常量2>)]...]
功能:将新元组插入指定表中
```

数据修改

```
UPDATE <表名>
SET <列名>=<表达式>[<列名>=<表达式>]
[WHERE <条件>]
功能:修改指定表中的满足WHERE字句条件的元组
SET字句给出<表达式>的值用于取代相应的属性列
```

三种修改方式

- 修改一个元组的值

```
UPDATE Student
SET id=1847118
WHERE id=20171847118
```

- 修改多个元组的值
- 修改所有个元组的值

```
UPDATE Student
SET age=age+1
```

练习: 1.修该操作系统课程的同学学分为5

```
UPDATE Course
SET Ccredit=5
WHERE Cname='OS'
```

2.将所有成绩清空

```
UPDATE SC
SET grade=0
```

1.实体完整性 2.主码不允许修改

删除数据

```
DELETE
FROM <表名>
WHERE <条件>
```

- 删除一个元组的值
- 删除多个元组的值
- 删除所有元组的值

查询数据

语句格式：

```
SELECT [ALL|DISTINCT] <目标列表达式>[<目标列表达式>]
FROM <表名或视图名>[<表名或视图名>] (SELECT语句)[AS]<别名>
[where<条件表达式>]
[GROUP BY<列名1>[HAVING<条件表达式>]]
[ORDER BY<列名2>[ASC|DESC]]
```

$$ \begin{align*} & Ssd & \end{align*} $$ 练习：查询product表中所有商品的名称和价格

```
SELECT prod_id,price FROM product
```

查询product表中所有供应商的id

```
SELECT vend_id FROM product
```

关系数据库简介

关系数据库理论

- 关系模型理论
- 关系规范化理论

关系数据形式化定义

```
单一的数据结构--关系
逻辑结构--二维表
建立在集合代数的基础之上
```

域(Domain)

域是一组具有相同数据类型的值的集合. 笛卡尔积.

- 元组: 笛卡尔积中每一个元素
- 分量:笛卡尔积中每一个元素的一个值
- 基数:一个域所允许的不同取值的个数

关系

笛卡尔积的一个子集

- 基本关系:实际存在的表
- 列是同质的(每一列的取值来自同一个域)
- 不同的列可能出资同一个域
- 列的顺序无所谓,列的次序可以任意交换
- 任意两个元组的候选码不能相同
- 行的数学无所谓,行的次序可以任意交换
- 分量必须取原子值
- 查询表
- 视图表:只有一个定义

码

- 候选码:一个极小的属性集合,去掉任意一个属性以后就不在称为一个候选码
- 全码:最极端的情况,关系模式的所有属性组是这个关系模式的候选码
- 主码:若一个关系有多个候选码,则选一个为主码
- 主属性:候选码的诸属性称为主属性,其他的称为非主属性
- 超码:候选码加上一个属性

关系的操作

- 常用的关系操作
- 查询操作:选择,投影,连接,除,并,差,交,笛卡尔积
  - 选择,投影,连接,并,差,笛卡尔积是五种基本操作
- 数据更新:插入,删除,修改
- 关系操作的特点:集合操作方式

关系的三类完整性操作

- 实体完整性:若属性A是基本关系的主属性,则属性A不能取空值.
- 实体完整性是针对基本关系而言的,一个基本表通常对应显示现实中的一个实体集
- 现实中的实体是可区分的,即他们具有某种唯一性标识
- 关系模型中以主码作为唯一标识
- 主码中的属性即主属性不能取空值
- 参照完整性:(F是基本关系R的一个或一组属性,但是不是关系R的码,K_S是基本关系S的主码.如果F与K_S对应则F是S 的外码,并称基本关系R为参照关系,基本关系S为被参照关系或目标关系)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
- 或者取空值
- 或者等于S中某个元组的主码值
- 空值
- 非空值
- 用户定义的完整性: 针对某一具体关系数据库的约束条件.

关系代数

关系代数是一种抽象的查询语言,它用对关系的运算来表达查询

运算的对象和结果都是关系

运算符	含义
$\cup$	并
-	差
$\cap$	交
×	笛卡尔集
σ	选择
π	投影
∞	连接
÷	除

- 查询

查询性别为男且年龄小于20岁的学生信息: $\sigma_{sex=男\wedge Sage<20}(Student)$

查询选修了2号课程的学生的学号,课程号,成绩: $\sigma_{Cno='2'}(SC)$

查询先行课为5的课程信息: $\sigma_{Cno='5'}(Course)$

- 投影 $\pi_{cno,course}(Course)$
- 连接 连接也吃呢个为θ连接,它是从两个关系相同的笛卡尔积中选取属性间满足一定条件的元组.
- 等值连接 θ为"="的连接运算称为等值连接 $$ R\infty_{A\theta B}S={{t_r}^\smallfrown {t_s}\ ｜t_r \in R \wedge

t_S \in S \wedge t_r[A]=t_s[B]} $$

- 自然连接 是一种特殊的等值连接 $$ R\infty_{A\theta B}S={{t_r}^\smallfrown {t_s}\ [U-B]｜t_r \in R \wedge t_S \in S \wedge t_r[A]=t_s[B]} $$
- 外连接 结果关系中也保存悬浮元组,其他属性上填空值
  - 左外连接
  - 右外连接

ex:1.查询所有选课学生的学号,姓名,性别,年龄,系,课程编号,成绩 $$ Strudent \infty SC $$ 2.查询所有被选修的课程编号,名称,先行课编号,学分,选修课学生编号,成绩 $$ Course \infty SC $$ 3.查询所有选课学生的学号,姓名,课程编号,成绩 $$ \pi_{sname,sno,cno,source}(Student \infty SC) $$ 4.查询选修了8号课程的学生的学号,姓名,系 $$ \pi_{sno,sname,sde}(\sigma_{cno=8}(SC) \infty Student) $$ 5.选了OS课程的学生的学号,姓名,选修课程名,成绩,有的学生可能不只选了OS $$ \pi_{sno,sname,cname,grade}(\pi_{cno='OS'}(Student \infty SC) \infty Studnet \infty SC \infty Course)) $$ 6.OS课程的成绩单,包括学生的学号,姓名成绩 $$ \pi_{sno,snme,grade}(\sigma_{cno='OS'}(Student \infty SC \infty Course)) $$ 7.所有学生的学号,姓名,系,选修课程号,成绩,有的学生可能没选任何课 $$ \pi_{sno,sname,sde,cno,grade}(Studnet 外\infty SC) $$ 8."DB"课程的先修课的课程编号,课程名和学分 $$ Course $$

- 除运算÷ 设关系R除以关系S的结果为关系T,则T包含在所有在R但不在S中的属性及其值,且T的元组与S 的元组的所有组合都在R中 $$ R \div S = {t_r[X]|t_r \in R \wedge \pi_Y(S)\subseteq Y_x}$$
  - 象集(image set) Z_x 给定一个关系 $R(X,Z)$,X,Z都是属性组。

    当 $t[x]=x$ 时，x喜爱R中的象集，为 $$Z_x={t[Z] | t \in R \quad and \quad t[x]=x}\ Y_x：x在R中的象集，x=t_r[X] $$

综合练习： 1.‘IS’系学生学生以及年龄>20岁学生的信息 $$\sigma_{Sdept=IS}（Student）\bigcup \sigma_{Sage>20}（Student）$$ 2.选修了‘0001’号学生所选的全部课程学生的学号 $$(\pi_{sno,cno}(SC) \div (\pi_{cno}(\sigma_{sno=0001}(Student))(SC)))\infty Student$$ 3.没选’c01‘课程学生的学号，姓名 $$(\pi_{sno}(Student)-\pi_{sno}(\sigma_{cno='c01'}(SC)))\infty \pi_{sno,sname}(Studnet)$$ 4.只选了’c01’课程的学生的学号 $$\pi_{sno}(\sigma_{cno=c01}(SC))-\sigma(SC \infty *{sc.sno=sc'.sno}*\rho{sc'}SC)$$

其他关系代数运算

- 关系更名
- 关系赋值
- 广义投影
- 聚集函数

空值处理

空值不等于任何值。