

Homework #1

1、给定下面的关系：图书（图书号，书名，作者，单价，库存量），读者（读者号，姓名，工作单位，地址），借阅（图书号，读者号，借期，还期，备注）。

使用关系代数表达式实现下列 1—5 小题：

(1) 检索读者 Rose 的工作单位和地址；

$$\pi_{\text{工作单位,地址}}(\sigma_{\text{姓名}='Rose'}(\text{读者}))$$

(2) 检索读者 Rose 所借阅过的图书的图书名和借期；

$$\pi_{\text{书名,借期}}(\sigma_{\text{姓名}='Rose'}(\text{图书} \bowtie \text{读者} \bowtie \text{借阅}))$$

(3) 检索未借阅图书的读者姓名；

$$\pi_{\text{姓名}}((\pi_{\text{读者号}}(\text{读者}) - \pi_{\text{读者号}}(\text{借阅})) \bowtie \text{读者})$$

(4) 检索 Ullman 所写的书的书名和单价；

$$\pi_{\text{书名,单价}}(\sigma_{\text{作者}='Ullman'}(\text{图书}))$$

(5) 检索借阅图书数目超过 3 本的读者姓名。

$$\pi_{\text{姓名}}(\sigma_{\text{book_count}>3}(\gamma_{\text{读者号},\text{COUNT(图书号)}\rightarrow\text{book_count}}(\text{借阅}) \bowtie \text{读者}))$$

注: γ 为分组操作符。

用 SQL 语言完成 6—10 小题：

(6) 检索 Ullman 所写的书的书名和单价；

```
SELECT 书名,单价
FROM 图书
WHERE 作者='Ullman';
```

(7) 检索读者“李林”借阅未还的图书的图书号和书名；

```
SELECT 图书号,书名
FROM 图书,读者,借阅
WHERE 图书.图书号=借阅.图书号 AND 读者.读者号=借阅.读者号 AND 姓名='李林'
AND 还期 IS NULL;
```

(8) 检索借阅图书数目超过 3 本的读者姓名；

```
SELECT 书名
FROM 读者,借阅
WHERE 读者.读者号=借阅.读者号
GROUP BY 读者号
HAVING COUNT(*)>3;
```

(9) 检索没有借阅读者“李林”所借的任何一本书的读者姓名和读者号；

```
SELECT 读者号,姓名
FROM 读者
WHERE 读者号 NOT IN
(
    SELECT 读者号
    FROM 读者,借阅
    WHERE 读者.读者号=借阅.读者号 AND 借阅.图书号 IN
    (
        SELECT 图书号
        FROM 借阅,读者
        WHERE 读者.读者号=借阅.读者号 AND 读者.姓名='李林'
    )
);
```

注：容易忽略从未借过书的读者

(10) 检索书名中包含“Oracle”的图书书名及图书号。

```
SELECT 图书号,书名
FROM 图书
WHERE 书名 LIKE '%Oracle%';
```

2、给定关系模式 $R(A,B)$ 、 $S(B,C)$ 和 $T(C,D)$ ，已知有下面的关系代数表达式，其中 p 是涉及属性 $R.A$ 的谓词， q 是涉及 $R.B$ 的谓词， m 是涉及 $S.C$ 的谓词。请写出与此关系代数表达式对应的 SQL 查询语句：

$$\pi_{(D)}[\sigma_{(p \wedge q \wedge m)}(R \bowtie S) \bowtie T]$$

```
SELECT D
FROM R,S,T
WHERE R.B=S.B AND S.C=T.C AND p AND q AND m;
```

3、已知有关系模式 $R(A,B,C,D,E)$ ， R 上的一个函数依赖集 $F=\{A \rightarrow BC, B \rightarrow CE, A \rightarrow B, AB \rightarrow C, AC \rightarrow DE\}$ 。

(1) 求 R 上 F 的一个最小函数依赖集（要求写出求解过程）；

将右边写出单属性并去除重复 FD:

$$\begin{aligned} F &= \{A \rightarrow B, A \rightarrow C, B \rightarrow C, B \rightarrow E, A \rightarrow B, AB \rightarrow C, AC \rightarrow D, AC \rightarrow E\} \\ &= \{A \rightarrow B, A \rightarrow C, B \rightarrow C, B \rightarrow E, AB \rightarrow C, AC \rightarrow D, AC \rightarrow E\} \end{aligned}$$

消去左部冗余属性:

由 $A \rightarrow C$ 可得 $AB \rightarrow C$ 是冗余属性;

由 $A \rightarrow C, AC \rightarrow D$ 可推出 $A \rightarrow D$, 因此可去除 $AC \rightarrow D$ 中的 C ;

由 $A \rightarrow C, AC \rightarrow E$ 可推出 $A \rightarrow E$, 因此可去除 $AC \rightarrow E$ 中的 C ;

$$F=\{A\rightarrow B, A\rightarrow C, B\rightarrow C, B\rightarrow E, A\rightarrow D, A\rightarrow E\}$$

消去冗余函数依赖:

由 $A\rightarrow B, B\rightarrow C$ 可推出 $A\rightarrow C$, 所以 $A\rightarrow C$ 是冗余属性;

由 $A\rightarrow B, B\rightarrow E$ 可推出 $A\rightarrow E$, 所以 $A\rightarrow E$ 是冗余属性;

$$F=\{A\rightarrow B, B\rightarrow C, B\rightarrow E, A\rightarrow D\}$$

(2) 求 R 的候选码, 并给出证明。

由 $A\rightarrow B, B\rightarrow C$ 可推出 $A\rightarrow C$;

由 $A\rightarrow B, B\rightarrow E$ 可推出 $A\rightarrow E$;

所以 $A\rightarrow\{A, B, C, D, E\}$, 候选码是 A.

Homework #2

1、已知有关系模式 $R(A, B, C, D, E)$, R 上的一个函数依赖集如下:

$$F=\{A\rightarrow BC, B\rightarrow CE, A\rightarrow B, AB\rightarrow C, AC\rightarrow DE, E\rightarrow A\}$$

(1) 求出 F 的最小函数依赖集:

将右边写出单属性并去除重复 FD:

$$F=\{A\rightarrow B, A\rightarrow C, B\rightarrow C, B\rightarrow E, A\rightarrow B, AB\rightarrow C, AC\rightarrow D, AC\rightarrow E, E\rightarrow A\}$$

$$=\{A\rightarrow B, A\rightarrow C, B\rightarrow C, B\rightarrow E, AB\rightarrow C, AC\rightarrow D, AC\rightarrow E, E\rightarrow A\}$$

消去左部冗余属性:

由 $A\rightarrow C$ 可得 $AB\rightarrow C$ 是冗余属性;

由 $A\rightarrow C, AC\rightarrow D$ 可推出 $A\rightarrow D$, 因此可去除 $AC\rightarrow D$ 中的 C;

由 $A\rightarrow C, AC\rightarrow E$ 可推出 $A\rightarrow E$, 因此可去除 $AC\rightarrow E$ 中的 C;

$$F=\{A\rightarrow B, A\rightarrow C, B\rightarrow C, B\rightarrow E, A\rightarrow D, A\rightarrow E, E\rightarrow A\}$$

消去冗余函数依赖:

由 $A\rightarrow B, B\rightarrow C$ 可推出 $A\rightarrow C$, 所以 $A\rightarrow C$ 是冗余属性;

由 $A\rightarrow B, B\rightarrow E$ 可推出 $A\rightarrow E$, 所以 $A\rightarrow E$ 是冗余属性;

$$F=\{A\rightarrow B, B\rightarrow C, B\rightarrow E, A\rightarrow D, E\rightarrow A\}$$

(2) 求 R 的候选码;

由 $A\rightarrow B, B\rightarrow C$ 可推出 $A\rightarrow C$;

由 $A\rightarrow B, B\rightarrow E$ 可推出 $A\rightarrow E$;

所以 $A\rightarrow\{A, B, C, D, E\}$, 又因为 $E\rightarrow A, B\rightarrow E$, 所以候选码是 A, B 和 E.

(3) R 属于第几范式? 为什么?

答: R 属于 BCNF, 因为 F^+ 的所有不平凡 FD 的左部都是超码.

(4) 请将 R 无损连接并且保持函数依赖地分解到 3NF。

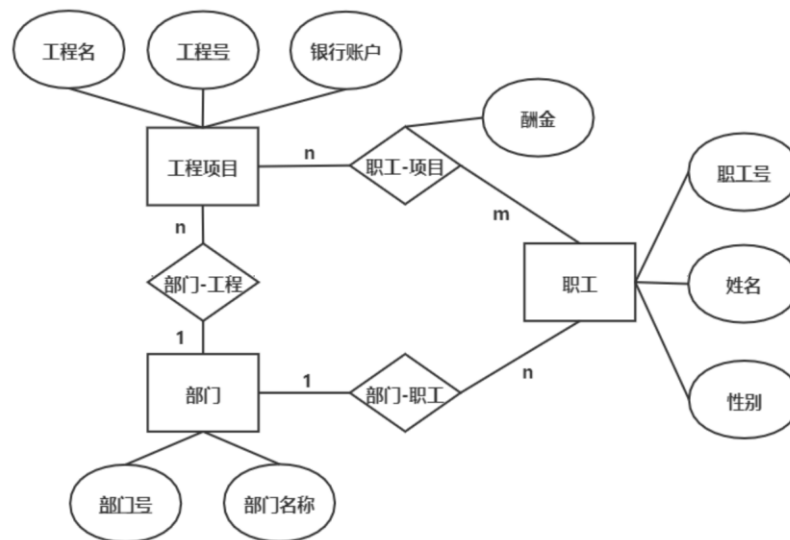
答: 已经是 BCNF, 无需分解.

2、假设某公司要开发一个信息管理系统。根据调研，获得该公司的业务规则如下(所有实体的码可以按日常生活中的一般规则处理)：

- (1) 公司下设几个部门，如技术部、财务部、市场部等；
- (2) 每个部门承担多个工程项目，每个工程项目属于一个部门。每个部门为其承担的每个项目分别开设独立的银行账户。
- (3) 每个部门有多名职工，每一名职工只能属于一个部门。
- (4) 一个职工可能参与多个工程项目，且每个工程项目有多名职工参与施工。根据职工在工程项目中完成的情况发放酬金。
- (5) 工程项目有工程号、工程名两个属性；部门有部门号、部门名称两个属性；职工有职工号、姓名、性别属性。

根据以上描述，请画出相应的 ER 模型（使用传统的 ER 建模符号），并将它转换为关系数据库模式。注：关系模式和属性名称均使用中文名称。

ER 模型:



关系数据库模式:

1. 部门(部门号,部门名称)
2. 工程项目(工程号,工程名,部门号,银行账户)
3. 职工(职工号,姓名,性别,部门号)
4. 职工-工程项目(职工号,工程号,酬金)

Homework #3

1、假设某磁盘具有以下特性：

- (1) 有 8 个盘面和 8192 个柱面；
- (2) 盘面直径为 3.5 英寸，其中内圈不存储数据，内圈直径为 1.5 英寸；
- (3) 每磁道平均有 256 个扇区，每个扇区 512 字节；
- (4) 每个磁道 10% 被用于间隙；
- (5) 磁盘转速为 7200 RPM；
- (6) 磁头启动到停止需要 1ms，每移动 500 个柱面另加 1ms。

回答下列问题：

(1) 磁盘容量是多少？

$$8192 * 8 * 256 * 512 \text{ 字节} = 8\text{GB}$$

(2) 如果所有的磁道拥有相同的扇区数，那么最内圈的磁道的位密度是多少？

$$\text{每个磁道容量 } N = 256 * 512 * 8 \text{ bits}$$

$$\text{最内圈扇区所占长度 } l = 1.5 * \pi * 90\% \text{ inch}$$

$$\text{因此最内圈磁道位密度 } N / l = 247238.5979 \text{ bpi}$$

(3) 如果一个块是 8KB，那么一个块的传输时间是多少？

8KB 为 16 个扇区，因此磁头需要经过 16 个扇区和扇区之间的 15 个间隙，覆盖的总长度占磁道的比值为： $(16 / 256) * 90\% + (15 / 256) * 10\% = 0.062109$

$$\text{磁盘转一圈所需时间为: } 1 / (7200 \text{ RPM}) = 8.333 \text{ ms}$$

$$\text{因此一个块的传输时间为: } 0.062109 * 8.333 \text{ ms} = 0.518 \text{ ms}$$

(4) 平均寻道时间是多少？

线段上任取两点，两点间距离的长度期望是线段长度的 1/3

$$\text{平均寻道数} = 8192 / 3 = 2730$$

$$\text{平均寻道时间} = 1 + 2730 / 500 = 6.46 \text{ ms}$$

(5) 平均旋转等待时间是多少？

$$\text{磁盘转一圈所需时间为: } 1 / (7200 \text{ RPM}) = 8.333 \text{ ms}$$

$$\text{平均旋转等待时间为旋转半圈所需的时间: } 8.333 / 2 = 4.17 \text{ ms}$$

2、假设某块磁盘的参数如下：容量为 36.7GB，传输速率为 45MB/s，旋转一圈的时间为 4ms，平均寻道时间为 5ms，最小寻道时间为 0.65ms（指磁头寻道到相邻磁道的时间），一个磁道大小为 180KB。如果磁盘块大小为 4KB，请回答下面问题（所有结果均四舍五入保留小数点后两位）：

(1) 随机读取 1000 个磁盘块需要多少时间(ms)？

$$\text{由题意知平均寻道时间 } t_1 = 5 \text{ ms, 平均旋转时间 } t_2 = 4 \text{ ms} / 2 = 2 \text{ ms}$$

$$\text{传输 1 个磁盘块时间 } t_3 = (4 \text{ KB} / 180 \text{ KB}) * 4 \text{ ms} = 0.08889 \text{ ms}$$

$$\text{因此随机读取 1000 个磁盘块所需时间为: } 1000 * (t_1 + t_2 + t_3) = 7088.89 \text{ ms}$$

(2) 假定(1)中的 1000 个磁盘块在单个磁道上连续存储，并且所有磁盘块存储在相邻的磁道上，此时读取这 1000 个磁盘块需要多少时间(ms)？

连续存储的 1000 个磁盘块占用的磁道数为： $\lceil 1000 * 4 / 180 \rceil = 23$ ，因此需要 22 次相邻磁道的最小寻道时间 t_4 。由于磁盘块连续存储，只考虑一次平均寻道时间 t_1 和 1 次平均旋转时间 t_2 ，因此读取这 1000 个磁盘块所需时间为： $t_1 + t_2 + 1000 * t_3 + 22 * t_4 = 110.19 \text{ ms}$