

## 16.5

因为有交换律, 所以先join哪个都是可以的。比如先r1自然连接r2, 最多产生1000行的关系, 因为C是r2的key。然后自然连接r3, 因为连接的E是r3的key, 所以最后也是最多1000行。

有效策略:

在关系r2的属性C和r3的E上建立一个索引。遍历r1, 然后根据索引从r2中寻找对应的C, 然后根据r2对应的那一行, 从r3中寻找对应的E。

## 16.6

关系大小的估算可通过计算第二个关系中每个元组的平均连接元组数来确定。对于r1中的每个元组, 平均会有  $1500/V(C,r2)=15/11$  个 r2的元组与之连接。中间关系将有  $15000/11$  个元组。将该中间关系与r3连接后, 结果约为10,227个元组 ( $15000/11 \times 750/100=10227$ )。高效策略应优先连接r1和r2, 因为中间关系的大小与r1或r2相近, 然后再将r3与此结果连接。

## 16.16

1. 利用 (dept\_name, building) 的B+树索引:

- 定位第一个满足 `dept_name = "Music"` 的元组。
- 通过指针链顺序检索后续元组, 直到 `building` 不再小于 "Watson"或 `dept_name` 不再等于 `Music`。

2. 过滤不满足条件的元组:

- 在检索到的元组中, 进一步筛选掉不满足 `budget < 55000` 的元组。

## 16.20

1. 直方图结构:

- 假设直方图H将属性A的值划分为多个范围 $r_1, r_2, \dots, r_n$ 。
- 每个范围 $r_i$ 包含:
  - 最小值 $r_i : low$
  - 最大值 $r_i : high$
  - 该范围内的元组数量 $H(r_i)$

2. 估计逻辑:

- **情况1:** 如果 $r_i : high \leq v$ (整个范围满足 $A \leq v$ ), 则直接累加该范围的元组数:  
估计值 $+= H(r_i)$
- **情况2:** 如果 $v$ 落在某个范围内(即 $r_i : low \leq v < r_i : high$ ), 假设该范围内值均匀分布, 按比例计算:  
估计值 $+= H(r_i) \times (v - r_i : low) / (r_i : high - r_i : low)$
- **情况3:** 如果 $v < r_i : low$ , 则忽略该范围(不满足条件)。

3. 最终结果:

- 将所有满足条件的范围的贡献值相加, 得到估计的元组数量。