

数据库系统原理第五次作业

2154312 郑博远

16.5 Consider the relations $r_1 = (A, B, C)$, $r_2 = (C, D, E)$, and $r_3 = (E, F)$, with primary keys A , C , and E , respectively. Assume that r_1 has 1000 tuples, r_2 has 1500 tuples, and r_3 has 750 tuples. Estimate the size of $r_1 \bowtie r_2 \bowtie r_3$, and give an efficient strategy for computing the join.

Solution:

The size of $r_1 \bowtie r_2 \bowtie r_3$ does not differ according to the computing order of the join operation, so consider the strategy of $(r_1 \bowtie r_2) \bowtie r_3$. Joining r_1 and r_2 will yield a result with a size of no more than 1000 tuples, since C is the primary key of r_2 , i.e. the C value is unique for each tuple in r_2 . Similarly, joining the result with r_3 will yield a result with a size of no more than 1000 tuples, since E is the primary key of r_3 . In conclusion, the size of $r_1 \bowtie r_2 \bowtie r_3$ is no more than 1000 tuples.

To compute the join efficiently:

- *Create an index on attribute C for r_2 , and one on attribute E for r_3*
- *For each tuple in r_1 , use the index of C in r_2 to find the only tuple with the same value of C in r_2 . Then, use the index of E in r_3 to find the only tuple with the same value of E in r_3 .*