Teacher: Matteo Lissandrini – Exercise Sheet: Query Optimization

Thanks to: Katja Hose, Gabriela Montoya



## 1 Join Ordering and Cardinality Estimation

## 1.1 Exercise

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  $r_1 \bowtie r_2 \bowtie r_3$  with a proper join ordering.

**Solution.** The relation resulting from the join of  $r_1, r_2$ , and  $r_3$  will be the same no matter which way we join them, due to the associative and commutative properties of joins.

So we will consider the size based on the strategy of  $((r_1 \bowtie r_2) \bowtie r_3)$ . This strategy follows the foreign key direction  $A \mapsto B \mapsto C$ .

Joining  $r_1$  with  $r_2$  will yield a relation of at most 1000 tuples, since C is a key for  $r_2$ . Likewise, joining that result with  $r_3$  will yield a relation of at most 1000 tuples because E is a key for  $r_3$ .

Therefore, the final relation will have at most 1000 tuples.

## 1.2 Exercise

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.

Assume that there are no primary keys, except the entire schema. Assume

- $|Values(C, r_1)| = 900$ ,
- $|Values(C, r_2)| = 1100$ ,
- $|Values(E, r_2)| = 50$ , and
- $|Values(E, r_3)| = 100$ .

Estimate the size of  $r_1 \bowtie r_2 \bowtie r_3$  and give an efficient strategy for computing the join.

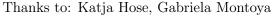
## Solution.

The estimated size of the relation can be determined by calculating the average number of tuples which would be joined with each tuple of the second relation.

In this case, for each tuple in  $r_1$ ,  $\frac{1500}{|Values(C,r_2)|} = \frac{15}{11}$  tuples (on average) of  $r_2$  would join with it.

Hence, the intermediate relation would have  $\frac{15 \times 1000}{11}$  tuples.

Teacher: Matteo Lissandrini – Exercise Sheet: Query Optimization





This relation is joined with  $r_3$  to yield a result of  $(\frac{15000}{11} \times \frac{750}{100} = 10227)$  tuples.

A good strategy should join  $r_1$  and  $r_2$  first, since the intermediate relation is about the same size as  $r_1$  or  $r_2$ .

Then  $r_3$  is joined to this result.