

Query Optimization | 2.C.iii

Entrance

This question is similar to [Relational Algebra | 8](#). There we had:

- $\min = 0$: nothing in R.B is the same with S.B (but this doesn't apply in our case because we have the foreign key constrain) \Rightarrow
- $\min = \min(|r| \times |s|)$: where both R.B and S.B ($V(B, r) = |r|$ and $V(B, s) = |s|$ | the maximum number of distinct values) are keys or if one of them is \Rightarrow
 1. $\min = |s|$: if R.B is the key $V(B, r) = |r|$
 2. $\min = |r|$: if S.B is the key $V(B, s) = |s|$
- $\max = n_r \times n_s$: everything in R.B is the same with S.B $\Rightarrow V(B, r) = 1$ and $V(B, s) = 1$ (the minimum number of distinct values).

Solution

Now we can clearly see that our last formula will depend on the:

1. $V(B, r)$ and $V(B, s)$.
2. foreign key constrain direction ($V(B, r) \geq V(B, s)$) $\Rightarrow \min(V(B, r), V(B, s)) = V(B, s)$

Here the exact value really depends in the structure and how the values of the records but we can give an average estimate in the size of the join when the records are very symmetric. Let us start with simpler case where R.B is a key $\Rightarrow \text{size}(r \bowtie s) = |s| \Rightarrow$

$$SC(B, s) = \frac{|s|}{V(B, s)}.$$

- On average each $V(B, s)$ distinct value it is approximately repeated $SC(B, s)$
- Because R.B is a key and R.S a foreign key \Rightarrow each $V(B, s)$ distinct value has 1 corresponding $V(B, r)$ distinct values. The $V(B, s)$ values not having a distinct value in $V(B, s)$ are not included in the join.

From these the $\text{size}(r \bowtie s) = V(B, s) \cdot SC(B, s) = |s|$ as expected. This is an example you should think about the arrows of the join:

r.B	s.B
1	1
2	1
3	2
4	2
5	3
6	3

- $V(B, s) = 3$
- $SC(B, s) = \frac{|s|}{V(B, s)} = \frac{6}{3} = 2$

- $size(r \bowtie s) = 3 * 2 = 6 = |s|$

This may seem pretty trivial and useless but it becomes important when R.B is not a key:

- $SC(B, s) = \frac{|s|}{V(B, s)}$
- $SC(B, r) = \frac{|r|}{V(B, r)}$
- Now for each $V(B, s)$ distinct value has $SC(B, r)$ corresponding $V(B, r)$ distinct values. Which implies we need to multiply $SC(B, r)$ with $SC(B, s)$ to get the number of records in the join for one value.

From these:

$$size(r \bowtie s) = V(B, s) \cdot (SC(B, r) \cdot SC(B, s)) = \frac{|s||r|}{V(B, r)}$$

example:

r.B	s.B
1	1
1	1
1	
2	2
2	2
2	
3	3
3	3
3	
4	
4	
4	

- $|s| = 6$
- $|r| = 12$
- $V(B, s) = 3$
- $V(B, r) = 4$
- $SC(B, s) = \frac{|s|}{V(B, s)} = \frac{6}{3} = 2$
- $SC(B, r) = \frac{|r|}{V(B, r)} = \frac{12}{4} = 3$
- $size(r \bowtie s) = \frac{|s||r|}{V(B, r)} = \frac{6*12}{4} = 18$

Overview

Generally speaking without including the foreign key constrain we have:

$$size(r \bowtie s) = \min(V(B, r), V(B, s)) \cdot (SC(B, r) \cdot SC(B, s)) = \frac{|s||r|}{V(B, r)}$$

Failures

--Trying to derive the exact formula

To analyze lets start with the same values in the table (we are going to use numbers for easier understanding). For the `min` case:

r.B	s.B
1	1
2	2
3	3
4	4
5	5
6	6

It doesn't matter if they are ordered. For simplicity of understanding we draw them in order. Now we are going to try to get to the max => make less distinct values. We can only start at $s.B$ because of the 2 point:

r.B	s.B
1	1
1	1
3	1
4	4
5	5
6	6

--her i realized that we can only give estimates

--here i was trying the other way

After this operation we get $size = (n_r - (V(B, r) - 1)) \times n_s$ For the next step there are 2 possible moves:

r.B	s.B	r.B	s.B
2	1	2	2
2	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1

For case 1: $size = (n_r - (V(B, r) - 2)) \times n_s$ For case 2:

$size = (n_r - (V(B, r) - 1)) \times (n_s - (V(B, s) - 1)) + 1$