

COMP9315 14s2 ... Q5

Based on primary keys, R has 1000 tuples and S has 2000 tuples

- (a) 50% of 1000 = 500
- (b) 2 (#Values R.b = 2)
- (c) 1 (aggregate has one result) (assume R.id was R.a)
- (d) 11211 is midpoint, tuples in range 113111-112111+1 = 1001
- (e) 0 (max value of S.y is 13345)
- (f) each S tuple joins 1 R tuple => 2000
- (g) select * from R left outer join S on (R.a=S.y)

Lower bound:

499 S.y values occur once, the other occurs 1501 times

=> the 499 S.y values each has one matching R tuple

=> the other S.y value also has one matcing R tuple

=> all the other R tuples (500) have R.a values that do not occur in S.y

=> $499 \times 1 + 1 \times 1501 + 2500 = 2500$ results

Upper bound:

each S.y value appears four times (2000/500, uniform)

=> every R tuple has 4 matching S tuples

=> $1000 \times 4 = 4000$ results