Date:
Name 1:
Name 2:

Assume we have two relations: R(a,b) with primary key a and S(a,c,d) with primary key (a,c) and foreign key constraint a references R.

We are given the following queries:

```
SELECT b, count(*) FROM
R
WHERE b >= 5
GROUP BY b
HAVING count(*) > 10;
```

```
SELECT a FROM
R NATURAL JOIN S
WHERE b>=5;
```

Assume:

- the values of b range from 0 to 9, and are equally distributed,
- most values of a are expected to appear in S

For each query:

1. Construct a parsing tree from each query (do the selection before the join)
2. For each parsing tree, construct a non-index query evaluation plan.
3. Approximately how much memory, in terms of the size of R and S do we need to be able to perform the query in on pass?
4. Assume that the size of the tuple of R is the same as the size of a tuple in S

You should make the following assumptions:

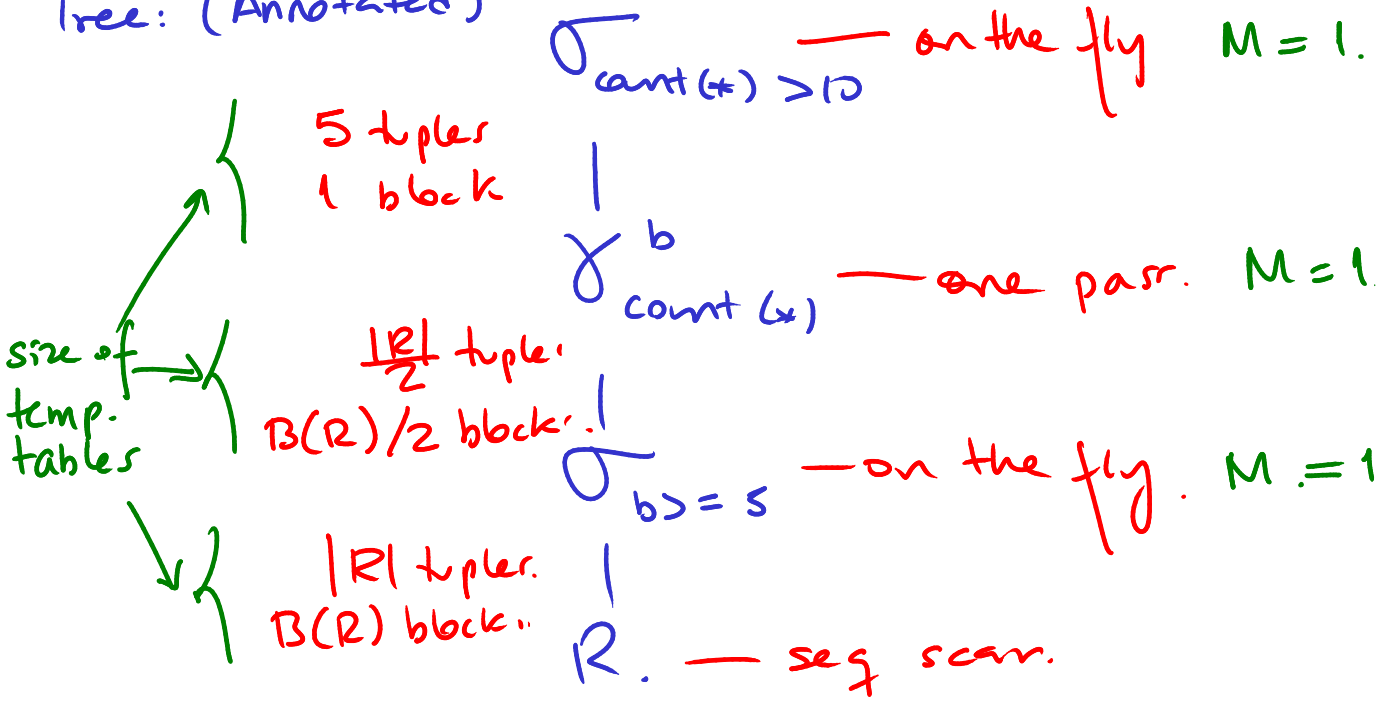
- All operations can be performed simultaneously (i.e. we need to allocate memory for each operation at the same time)
- Assume that B(S) > B(R)

```
SELECT b, count(*) FROM R
WHERE b>=5
GROUP BY b
HAVING count(*) > 10
```

Equivalent RA:

$$\sigma_{count(*)>10}\ \gamma^{b}_{count(*)}\ \sigma_{b>=5}\ R$$

Tree: (Annotated)

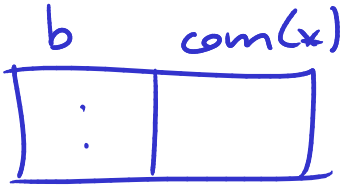


$\sigma_{b>=5}$ will output only 1/2 of R
because selectivity (b>=5) = 0.5

$\Rightarrow \gamma$ will only inspect 1/2 of tuples in R.

Memory required for γ:

Result will be



} # different b?

Selection $\sigma_{b>=5}$ feeds only b between 5 and 9. $|\gamma^b|$ = # diff values of b in input to γ.

$$|\gamma^b| = 5 \text{ tples} \quad (5, 6, 7, 8 \text{ and } 9)$$

$$\Rightarrow B(\gamma^b_{count(x)}) \cong 1 \text{ block.}$$

We can answer the query with 3 blocks and one pass.

Cost: $B(R)$

The expected size of the result is:

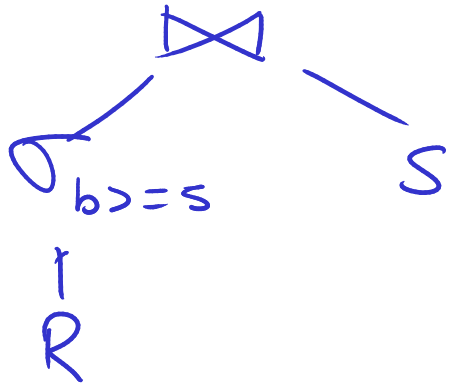
5 tples!!

```
SELECT * FROM
  R NATURAL JOIN S
  WHERE b >= 5
```

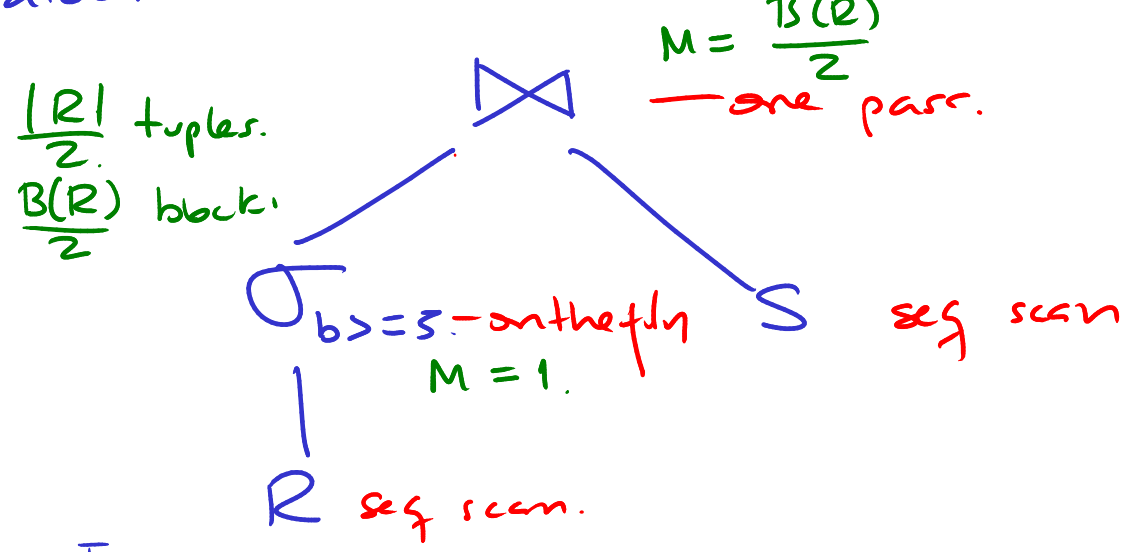
We do σ before $\bowtie$

$$(\sigma_{b>=5} R) \bowtie S$$

Tree.



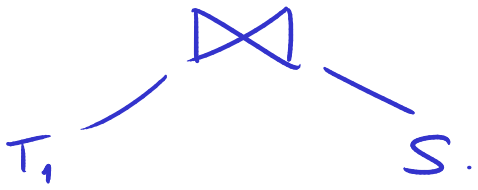
Annotated:



Selection generates a temp table T_1.

Selectivity $(b>=5) = 0.5$

$|T_1| = \frac{|R|}{2}$ $B(T_1) = \frac{B(R)}{2}$



$B(R.) < B(S). \Rightarrow B(T_1) < B(S).$

So we would need to load T_1 in memory.

$$M = B(T_1) = \frac{B(R)}{2}$$

So we can do the query one pass if $M > \frac{B(R)}{2}$

How many tuples in Result of query?

Because T_1 only has $\frac{1}{2}$ tuples in R

and S has FK on R(a)

Result will contain approximately $\frac{|S|}{2}$ tuple