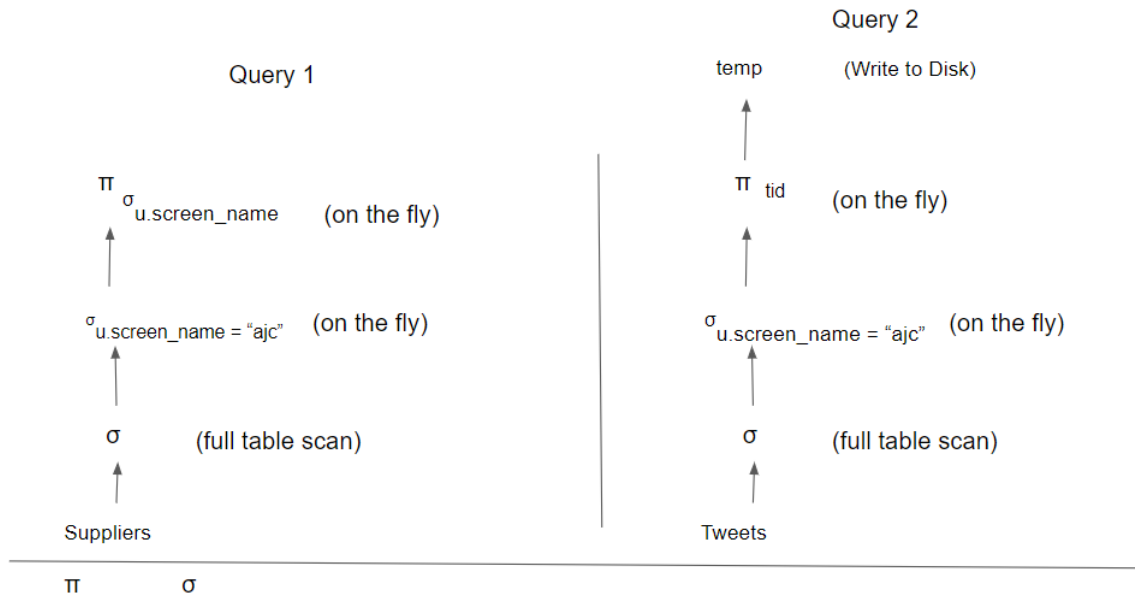


Author: Samuel Rettig

A)



Query 1:

Num of bytes per record: X

$$\begin{aligned} X &= 1(\text{bytes in char}) * 80(\text{max char amount for screen_name}) + \\ &1(\text{bytes in char}) * 80(\text{max char amount for user_name}) + \\ &1(\text{bytes in char}) * 80(\text{max char amount for category}) + \\ &1(\text{bytes in char}) * 80(\text{max char amount for sub_category}) + \\ &1(\text{bytes in char}) * 80(\text{max char amount for state}) + \\ &4(\text{bytes in int}) + 4(\text{bytes per int}) = X \\ X &= 408 \text{ bytes/record} \end{aligned}$$

Number of records per page = Y

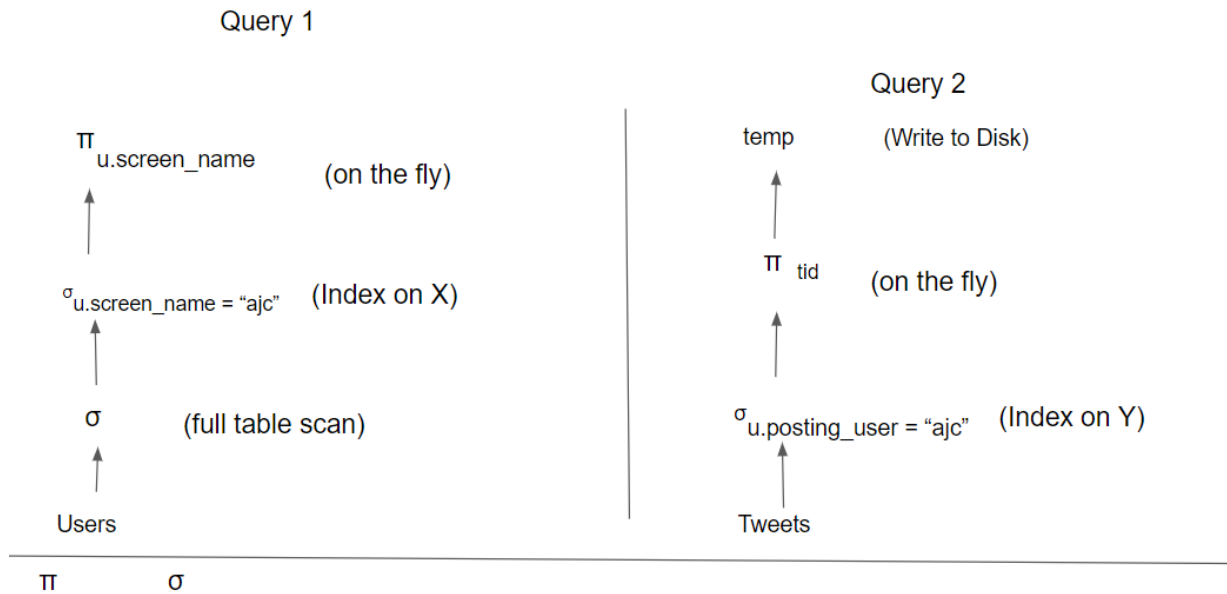
$$Y = \text{floor}(4000/X) = \text{floor}(4000 / 408) = 9 \text{ (9.8... if not floored)}$$

Number of pages = Z

$$Z = \text{ceil}(5000 / Y) = (5000 / 9) = 556 \text{ pages}$$

Since we are not using the header page, the final result is 556

Disk I/O = 556



X is any index on User that contains screen_name as the only/first element

Y is any index on Tweets that contains posting_user as the only/first element

Query 2:

Disk IO Cost = Disk I/O cost of full scan + Disk I/O cost of rows

Full table:

Number of bytes per record is X

$X = 8$ (bytes (bigint) for tid) + 4 (bytes for post_day) +

$$4(\text{bytes for post_month}) + (4 \text{ bytes for post_year}) + 4 (\text{bytes for rwtweetCT}) +$$

$$1 * 80 (\text{ char posting user})$$

$$X = 104 \text{ bytes/record}$$

$$\text{Number of records per page} = Y$$

$$Y = \text{floor}(4000 / 104) = 38 \text{ records/page}$$

$$\text{Total number of pages} = Z$$

$$Z = \text{ceil}(10000 / 38) = 264 \text{ pages}$$

Disk I/O cost for full scan is 264

$$\text{Row cost} =$$

$$\text{Number of rows to write is} = X$$

$$X = 10000 (\text{amount of tuples}) * 0.1 (\text{selectivity factor}) = 1000$$

$$\text{Number of bytes per row} = Y$$

$$Y = 8 \text{ bytes/row}$$

$$\text{Number of rows per page} = Z$$

$$Z = \text{floor}(4000 / 8) = 500$$

$$Z = 500 \text{ rows/ page}$$

$$\text{Number of pages to write to: (following x,y,x,) } W$$

$$W = \text{ceil}(1000 / Z) = (1000 / 500) = 2 \text{ pages}$$

$$W = 2 \text{ pages}$$

Disk I/O cost is 2 pages

Final overall cost = 264 + 2 = 266