COMMON MISTAKES

- Included unnecessary projections in the relational algebra expressions.
- Failed to specify the algorithms used for each operation in the evaluation plan.
- Failed to sort the input relations of merge join on the join attributes
- Allocated memory blocks more than the available memory at one time.
- Failed to correctly estimate the size of joins.
- Failed to correctly estimate the size of disjunctive selection.
- Failed to correctly calculate the blocking factor for intermediate results.
- Failed to include the cost of writing intermediate results to disk for materialization.

INCOMPLETE SHORT ANSWERS

(a) [5%]

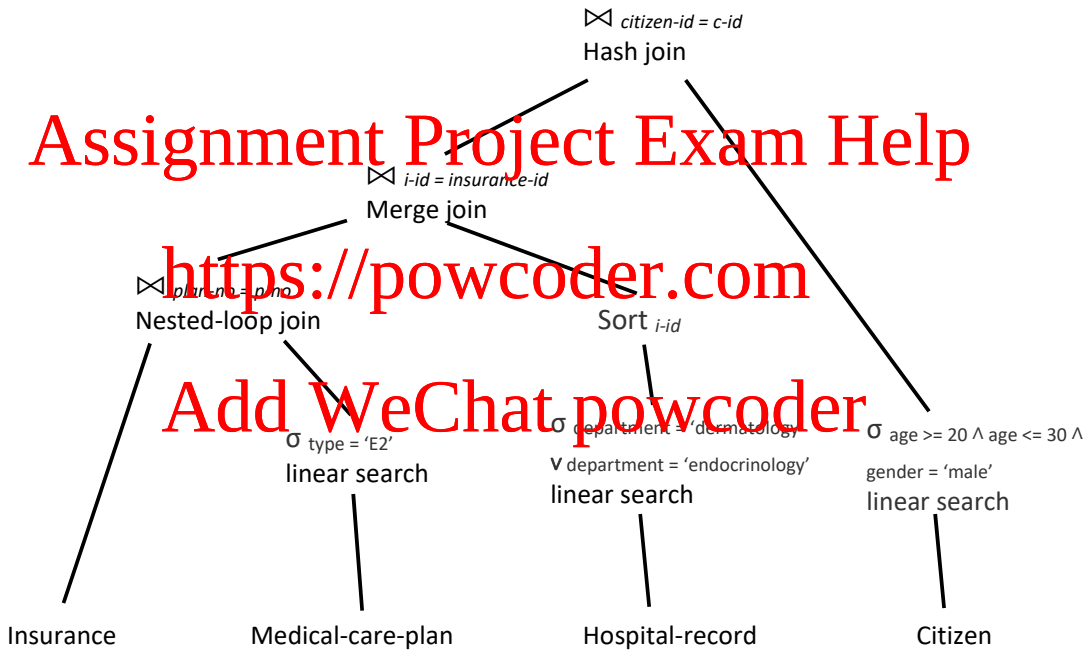
σ *age* <= 30 $\wedge$ *age* >= 20 $\wedge$ *gender*='male' $\wedge$ *type*='E2' $\wedge$ (*department* = 'dermatology' *or* 'endocrinology') (
((*Citizen* $\bowtie$ *citizen-id = c-id Hospital-record*)
$\bowtie$ *i-id = insurance-id Insurance*) $\bowtie$ *plan-no = p-no Medical-care-plan*)

(b) [10%]

18000000 * 1/2 * 1/10 * (1- (1-1/59) * (1-1/60)) * 1/10
= 3000

(c) [65%]

One possible evaluation plan using left-deep join order and materialization:



Memory allocation

- Selection: 1 block for input and 59 blocks for output
- Nested-loop join: 1 block for each of the two inputs and 58 blocks for output
- Sort (merge stage): 10 blocks for each of the inputs and output
- Merge join: 25 blocks for each of the two inputs and 10 blocks for output
- Hash join (partitioning): 12 blocks for input and 1 block for each of the 48 partitions

Statistic estimation

- Number of tuples at the output of Selection of Citizen: 300,000
- Number of blocks at the output of Selection of Citizen: 4,000
- Number of tuples at the output of Selection of Hospital-record: 600,000

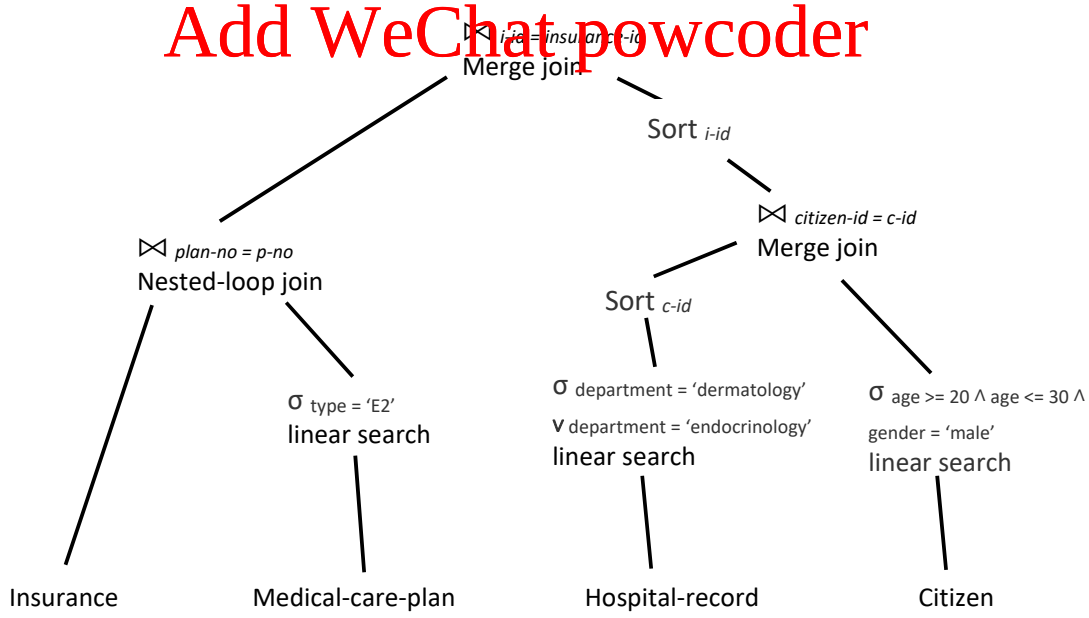
- Number of blocks at the output of Selection of Hospital-record: 6,000
- Number of tuples at the output of Selection of Medical-care-plan: 60
- Number of blocks at the output of Selection of Medical-care-plan: 1
- Number of tuples at the output of Nested-loop join: 900,000
- Number of blocks at the output of Nested-loop join: 27,273
- Number of tuples at the output of Merge join: 60,000
- Number of blocks at the output of Merge join: 2,400
- Number of tuples at the output of Hash join: 3,000
- Number of blocks at the output of Hash join: 167

Cost estimation (in number of block transfers (t_T) and number of disk seeks (t_S))

- Selection of Citizen: 84000 t_T + 136 t_S
- Selection of Hospital-record: 186000 t_T + 204 t_S
- Selection of Medical-care-plan: 11 t_T + 2 t_S
- Nested-loop join: (120000 + 1 + 27273) t_T + (942 + 1) t_S
- Sort: 48000 t_T + 3800 t_S
- Merge join: (27273 + 6000 + 2400) t_T + 1571 t_S
- Hash join: ((2400+4000)*3) t_T + ((2400/12+2400)+(4000/12+4000) + 48*2) t_S
- Total cost: 520158 t_T + 13686 t_S ~ 107 seconds

(d) [15%]

One possible better evaluation plan using any join orders:



(e) [5%]