

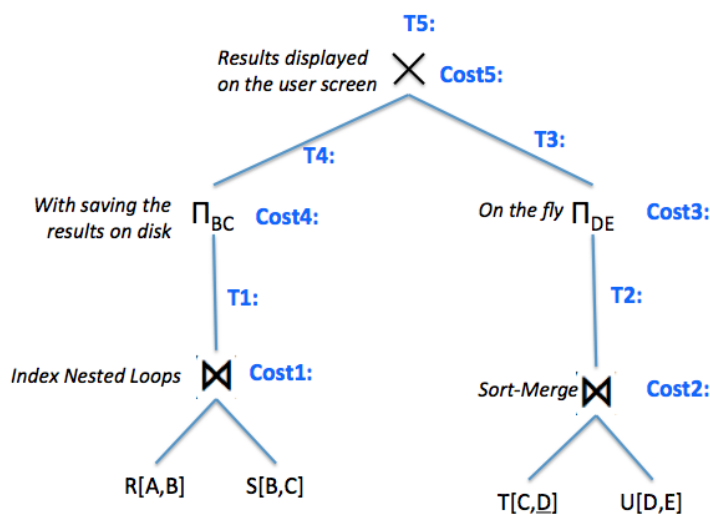
# Sept 2018

**Q. 1** (2 pts) Consider the 4 relations  $R[A, B]$ ,  $S[B, C]$ ,  $T[C, D]$ ,  $U[D, E]$  and you know the following information about each one of them:

$R[A, B]$	$S[B, C]$	$T[C, D]$	$U[D, E]$
has a B+tree on B	has $P_S$ pages	has $P_T$ pages	has $P_U$ pages
has $P_R$ pages	A tuple has size $T_s$	has $C_T$ number of tuples	has $C_U$ number of tuples
B has selectivity factor $f_B$	50% of B is null	T.D is a key	D has no null values
has $C_R$ number of tuples			

You also know that a page on the disk is of size P and no operator does duplicate elimination. Furthermore, the memory has 3 buffers, and each buffer is as big as a page of the disk. A B+tree lookup costs 3.

Based on that information write the formula that computes the cost of each operators (Cost1, Cost2, Cost3 and Cost4) and for each operator write the number of tuples that the operator generates (T1, T2, T3 and T4). In the formula you can use any of the values mentioned in the table above and any of the variables Cost1, Cost2, Cost3, Cost4, T1, T2, T3, and T4.



**Answer:**

$$\sim \text{Cost1} = P_S + |S| * \frac{50}{100} * (3 + f_B * |R|), \text{ (where } |S| = P_S * \lfloor \frac{P}{T_s} \rfloor \text{)}$$

$$T1 = |S| * \frac{50}{100} * f_B * |R|, \text{ (where } |S| = P_S * \lfloor \frac{P}{T_s} \rfloor \text{)}$$

$$\text{Cost2} = 2P_t \log P_T + 2P_u \log P_u + P_T + P_U,$$

$$T2 = C_u,$$

$$\text{Cost3} = 0,$$

$$T3 = T2 = C_u,$$

$$\text{Cost4} = \frac{T1}{\lfloor \frac{P}{t_s} \rfloor}, \text{ (Cost to save the results on the disk to be used by the } \times \text{)}$$

$$T4 = T1,$$

$$\text{Cost5} = \text{Cost4} * \frac{T3 * t_u}{B-1}, \text{ (where } T3 * t_u \text{ is the space of the } T3 \text{ tuples, and } B \text{ is the number of buffers available, and we put Cost4 because we need to read the } T4 \text{ tuples from the disk where we saved them)}$$

$$T5 = T3 * T4$$