BD2 – April 9th, 2018
Please feel free to answer this test in English, Italian, or any mixture

1. Consider a schema R(IdR, A, B, …, IdS*), S(IdS,…), T(IdT, C, D,…, IdS*) and the following query

```
SELECT DISTINCT T.C. count(*)
FROM        R, T
WHERE       R.IdS = T.IdS And R.A ≤ 100 And T.D = 10
GROUP BY    T.C, T.D
```

Assume that R is stored with a dense clustered index on A, while S and T are stored as heap files. Primary keys are R.IdR, S.IdS and T.IdT, while R.IdS and T.IdS are foreign keys that refers to S.
Assume that unclustered RID-sorted index are defined on all the primary and foreign keys, and on T.D.
Assume that the size of all indexes only depend on the number of RIDs, as indicated in the table below.
Assume that every page is 4.000 bytes long, and that every attribute uses 4 bytes. Assume a buffer size of 100 pages. If you need Cardenas formula Φ(n,k), approximate it with min(n,k).

	NReg	NPag	NLeaf of Indexes	NKey	Min	Max
R	500.000	10.000	1.000			
S	100.000	2.000	200			
T	10.000.000	50.000	20.000			
Idx.R.A			See R	100	0	1.000
Idx.T.D			See T	100	0	1.000

a) Is DISTINCT redundant?
b) Compute the selectivity factor of the three predicates in the condition
c) Compute the cost of accessing $\sigma_{A\leq 100}(R)$, $\sigma_{D\leq 10}(T)$ (useless for this query) and $\sigma_{D=10}(T)$, with and without index. Remember that the index on R.A is clustered.
d) Compute the cost of a MergeJoin plan that uses the cheapest plans for the selections (ignore the fact that IdS is not a key for neither table) (draw the plan first)
e) Compute the cost for an IndexNestedLoop plan where R is the outer (left-hand-side) leaf (draw the plan first)

2. a) Assume two relation R(a,b) and S(a,c). Define the left natural outer join(R,S) (leftjoin(R,S)) and give a very simple example.
   b) Write a two-lines (more or less) pseudo code specification for IndexNestedLoop. Briefly specify (using natural language or psudo-code) how one could generalize IndexNestedLoop to LeftIndexNestedLoop in order to compute left outer join
   c) Compare the cost of LeftIndexNestedLoop(OE,OI,condition)) with the cost of IndexNestedLoop(OE,OI),condition)

3. a) Explain (briefly) why heap organization is the one that is most commonly used
   b) In which situations the heap organization is the best one?
   c) In which situations is the hash (procedural) organization the best one?