

Examen DBD. 17 juny 2021. (1/2)

Cognoms..... Nom.....

Les notes es publicaran el dia 24 de juny. La revisió es farà online i s'ha de sol·licitar per correu electrònic al professor abans del dia 28.

Totes les preguntes tenen el mateix pes.

1) L'algoritme de *Hash Join* funciona, com sabeu, en dos passos: en un primer pas reorganitza les files de les dues taules fent servir una mateixa funció de *hash* que assigna cada fila a una de les p parts en què es parteix cada taula i, en un segon pas, fa la *join* part a part amb l'algoritme de *block nested loops*. Si suposem que tenim $M + 2$ pàgines de memòria, per garantir que el primer pas es pot fer amb una sola passada per taula cal que $M + 2 \geq p + 1$ de manera que tenim un *buffer* d'entrada i p *buffers* de sortida. D'altra banda, per garantir que en els *nested loops* del segon pas la taula petita càpiga a memòria cal que $B_{SMALL} / p \leq M$, atès que cada part de la taula més petita ocuparà B_{SMALL} / p blocs (recordeu que el cost de BNL és $B_R + B_S * (\lceil B_R / M \rceil)$). És satisfent simultàniament aquestes condicions que el cost de l'algoritme es correspon a llegir dos cops i escriure un cop cada taula i la conjunció d'aquestes condicions dóna lloc a la que ja coneixeu, $B_{SMALL} \leq M^2 + M$. Cas que aquesta condició no es compleixi, ens podem plantejar dos casos extrems:

- A. Triar p tan gran com es pugui garantint que el primer pas es fa amb una sola passada.
- B. Triar p tan petit com es pugui garantint que la taula petita càpiga a memòria en els *nested loops*.

Ompliu la taula següent amb els valors de p i el cost de cada pas en funció de B_{BIG} , B_{SMALL} i M . A la darrera columna indiqueu amb *S*, *N* o *?* si el cost total és sempre millor que el de fer directament la *join* amb BNL de les taules originals, sempre pitjor o no es pot assegurar.

	p	Cost pas 1	Cost pas 2	Millora BNL?
A				
B				

2) Com bé sabeu, quan hem estudiat els índexs hem suposat que a les entrades s'hi guardaven parelles (clau, adreça). Pensem com afectaria els costos d'un B+ que cada entrada de l'índex contingués una fila sencera de manera que ja no distingiríem entre índex i taula sinó que les dues coses s'integren en una.

Suposant que la mida d'un bloc són 12 KB, que les files ocupen 1 KB i que una entrada (clau, adreça) ocupa 0.05 KB, ompliu les taules següents.

ENTRADES (CLAU, ADREÇA)	Ordre arbre (d)	Cost una tupla	Cost diverses tuples (FS 0.1)	Cost diverses tuples (FS 0.9)
$ T = 10$				
$ T = 1000$				
$ T = 10000$				

ENTRADES FILA SENCERA	Ordre arbre (d)	Cost una tupla	Cost diverses tuples (FS 0.1)	Cost diverses tuples (FS 0.9)
$ T = 10$				
$ T = 1000$				
$ T = 10000$				

Examen DBD. 17 juny 2021. (2/2)

Cognoms..... Nom.....

Les notes es publicaran el dia 24 de juny. La revisió es farà online i s'ha de sol·licitar per correu electrònic al professor abans del dia 28.

Totes les preguntes tenen el mateix pes.

3) Com ja sabeu, una de les aplicacions de les vistes materialitzades és la de magatzems de resultats parcials de consultes multidimensionals gràcies a la capacitat dels SGBD de reescriure consultes expressades sobre les taules en consultes sobre les vistes que tenen un cost menor que les consultes originals. El cas més simple és aquell en què una consulta es reescriu fent servir una vista materialitzada, però també és possible reescriure una consulta fent servir més d'una vista. Donada la taula CentMilResp(ref, pobl, edat, cand, val) i les consultes següents, reescriuiu, si és possible, cada consulta en funció de les altres (una o més) o indiqueu "no és possible" quan escaigui.

(C1) SELECT cand, edat, MIN(val) AS mn FROM CentMilResp WHERE edat > 50 GROUP BY cand, edat;

(C2) SELECT cand, edat, COUNT(*) AS n, MAX(val) AS mx FROM CentMilResp WHERE edat > 40 GROUP BY cand, edat;

(C3) SELECT cand, edat, MAX(val) AS mx, MIN(val) AS mn FROM CentMilResp WHERE edat > 50 GROUP BY cand, edat;

(C4) SELECT cand, edat, COUNT(*) AS n, MAX(val) AS mx, MIN(val) AS mn FROM CentMilResp WHERE edat <= 60 GROUP BY cand, edat;

(C5) SELECT edat, COUNT(*) AS n FROM CentMilResp GROUP BY edat;

C1	
C2	
C3	
C4	
C5	