

Examen DBD. 27 juny 2018.

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

Les notes es publicaran el dia 4 de juliol. La revisió es farà el dia 6 de juliol a les 12h a la sala Omega S205.

Totes les preguntes tenen el mateix pes.

1) Considereu una taula `Menges(codi, nom, epoca, periode, pais, data)` **i aquestes estadístiques:**

`Ndist(nom) = 18.000`, `Ndist(epoca) = 24`, `Ndist(pais) = 200`, `min(data) = '01/01/2000'`, `max(data) = '31/12/2000'`,
`Ndist(periode) = 5`, `|Menges| = 20.000`, `BMenges = 4.000`

Suposem un índex B+ per l'atribut *epoca*, un B+ per *data*, un hash per *pais* i un cluster per *periode*. L'ordre dels arbres és 150 i el factor de càrrega és 2/3 per als B+ i clusters i 0.8 per al hash. Ens plantegem aquesta consulta:

```
SELECT * FROM Menges WHERE (data BETWEEN '01/07/2000' AND '31/12/2000' OR epoca IN ('Nadal', 'Sant Joan', 'Ferragosto')) AND pais IN ('Tabàrnia', 'Ítaca', 'Flandes', 'Góndor') AND periode <> 'dia'
```

Tingueu present l'algoritme de selecció per calcular les dades següents:

FS: |O|:.....

Blocs accedits índex data: Blocs accedits índex epoca: Blocs accedits índex pais:

Blocs accedits índex periode: Blocs accedits taula Menges:

2) Considereu les vistes següents definides a partir de la taula de l'exercici 1:

```
CREATE VIEW V1 AS SELECT * FROM Menges WHERE data < '01/04/2000'

CREATE VIEW V2 AS SELECT pais, COUNT(*) FROM Menges GROUP BY pais

CREATE VIEW V3 AS SELECT pais, COUNT(*) FROM V1 GROUP BY pais
```

Considereu ara aquestes actualitzacions:

A1: `UPDATE V2 SET pais = 'P1' WHERE pais = 'P2'`

A2: `DELETE FROM V2 WHERE pais = 'P'`

A3: `INSERT INTO V2 VALUES ('P', 0)`

Quines actualitzacions es poden a fer? Ompliu la taula següent amb S o N i justificant el Sí amb l'única operació propagable cap a la taula base i el No indicant "No Possible" o bé amb dues operacions propagables.

ACT	S/N	JUSTIFICACIÓ
A1		
A2		
A3		

Segueix ----->>>>

Feu ara el mateix per a la vista V3.

ACT	S/N	JUSTIFICACIÓ
A1		
A2		
A3		

3) Considereu la taula $T(\dots, a, b, \dots)$ i els índexos B (un B+), BM (un bitmap) i H (un hash), tots definits sobre l'atribut a . La taula té 1.000.000 tuples i ocupa 10.000 blocs. nbits val 50.000, l'ordre dels arbres és 75, el factor de càrrega dels arbres és $2/3$ i el del hash 0.8.

a) Per a cada índex, digueu si no té restricció pel que fa a nombre de repeticions per valor de l'atribut a o el màxim de repeticions que pot tolerar:

Índex B --> Índex BM --> Índex H -->

b) Donat un valor constant v , considereu la consulta

```
SELECT COUNT(*) FROM T WHERE a = v
```

Digueu quin índex és millor en funció del nombre de repeticions k . En cas d'empat, indiqueu tots els índexos empatats en primera posició

$k = 1$ --> $k = 10$ --> $k = 100$ --> $k = 1000$ -->
 $k = 10000$ --> $k = 100000$ --> $k = 1000000$ -->

c) Feu el mateix per a aquesta altra consulta:

```
SELECT SUM(b) FROM T WHERE a = v
```

$k = 1$ --> $k = 10$ --> $k = 100$ --> $k = 1000$ -->
 $k = 10000$ --> $k = 100000$ --> $k = 1000000$ -->

d) Hi ha alguna altra estructura que superi a les altres tres per a tots els valor de k en aquesta segona consulta?