

DISSENY CONCEPTUAL DE BASES DE DADES

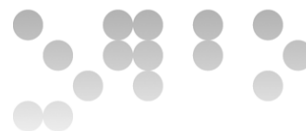
Presentació

Aquesta prova consta de 5 exercicis enfocats a determinar el grau de comprensió dels mòduls didàctics després de l'estudi dels mòduls 4 i 5.

Objectius i competències

La realització d'aquesta PAC ha de permetre demostrar que s'han assolit les competències següents:

- Conèixer l'estructura física que utilitza la base de dades per a emmagatzemar, de forma no volàtil, les dades.
- Fer el disseny físic de la base de dades a partir del disseny lògic, adaptat a les característiques d'un SGBD concret i definir els índexs necessaris i convenients a cada taula per tal que les aplicacions tinguin un bon rendiment quan accedeixen a la base de dades.
- Conèixer els mecanismes de processament i optimització de consultes, i així poder-les plantejar de la forma més eficient possible.
- Conèixer les diferents estratègies d'implementació de les operacions d'àlgebra relacional per a poder avaluar el cost de les consultes.
- Interpretar i optimitzar adequadament el pla d'execució d'una consulta.
- Presentar les vistes com a elements de disseny extern.
- Conèixer l'abast dels mecanismes de seguretat d'una base de dades.
- Conèixer el funcionament dels índex B+.



Descripció de la PAC

EXERCICI 1

A partir del següent disseny lògic:

Person (IPF, **name**, **surname**, birthDate)

Vehicle (plate, **IPF**, **brandName**, model, manufacturingYear)

{IPF} is foreign key of Person

Ens diuen que volen eliminar la restricció de la clau forana entre Vehicle i Person, i que es desitja que s'implementi aquesta restricció utilitzant disparadors.

Per mantenir la integritat referencial, quants disparadors serien necessaris? Indiqueu la funció que tindria cadascun d'ells, i si s'haurien d'executar 'after' o 'before'.

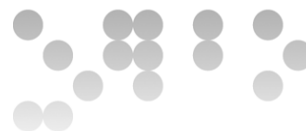
SOLUCIÓ:

Caldrien quatre disparadors. La funció d'aquests hauria de ser:

- InsertVehicle: Cal assegurar que existeix la Person que consta en el camp IPF.
- ModifyVehicle: Cal assegurar que, de canviar-se el camp IPF, existeixi a Person.
- DeletePerson: Cal assegurar que no existeixi un Vehicle amb el IPF de la Person.
- ModifyPerson: De canviar-se el camp IPF cal assegurar que no existeix un Vehicle amb aquest IPF.

No cal disparador per a les eliminacions de files a la taula Vehicle, ni per les insercions en la taula Person.

Tots els disparadors s'haurien d'executar abans de realitzar les operacions (before).



EXERCICI 2

Hem preparat la vista següent:

```
CREATE VIEW V_SubjectGroup2 AS
SELECT
    StudentGroup.subject,
    StudentGroup.code,
    Subject.name,
    Subject.credits,
    Studentgroup.Maxstudent
FROM Subject FULL OUTER JOIN StudentGroup ON
    Subject.Code = Studentgroup.Subject;
```

Sobre la qual fem aquesta consulta:

```
SELECT DISTINCT(name) FROM V_SubjectGroup2;
```

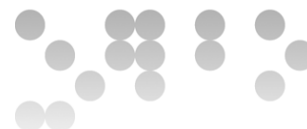
I obtenim el següent pla d'execució:

OPERATION	OBJECT_NAME	OPTIONS	COST
SELECT STATEMENT			7
HASH		UNIQUE	7
VIEW	VW_F01_0		6
HASH JOIN		FULL OUTER	6
Access Predicates			
SUBJECT.CODE=STUDENTGROUP.SUBJECT			
INDEX	PK_STUDENTGROUP	FAST FULL SCAN	2
TABLE ACCESS	SUBJECT	FULL	3

Com que sabem que tota assignatura té assignat, com a mínim, un grup, fem la següent consulta i ens trobem que retorna els mateixos resultats (files):

```
SELECT DISTINCT(name) FROM Subject;
```

En aquest cas el pla d'execució és el següent:



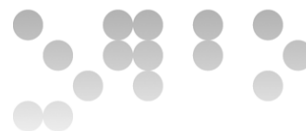
OPERATION	OBJECT_NAME	OPTIONS	COST
SELECT STATEMENT			4
HASH		UNIQUE	4
TABLE ACCESS	SUBJECT	FULL	3

Per què creus que el rendiment al primer pla d'execució és menor? Quin tipus d'optimització hauria d'aplicar l'SGBD a la primera consulta per a obtenir un cost similar al de la segona consulta?

SOLUCIÓ:

Les dues SELECTs retornen la mateixa informació: un camp que correspon a la taula Subject. En el cas de la SELECT sobre la vista l'optimitzador no detecta que no cal fer les combinacions amb totes les taules quan planifica la consulta, per les limitacions inherents al seu funcionament, independentment la consulta consti com a vista.

En el cas que l'Oracle fes una bona optimització semàntica (mòdul 5 pag 14), detectaria que la combinació és innecessària i el cost final seria l'equivalent al de la segona consulta.



EXERCICI 3

Amb la finalitat de fer una campanya telefònica informativa es desitja recuperar tots els telèfons dels clients de la base de dades que comencen per '93'.

```
SELECT * FROM Person WHERE phoneNumber LIKE '93%';
```

S'observa que la consulta té el cost següent:

OPERATION	OBJECT_NAME	OPTIONS	COST
SELECT STATEMENT			137
TABLE ACCESS	PERSON	FULL	137
Filter Predicates			
PHONENUMBER LIKE '93%'			

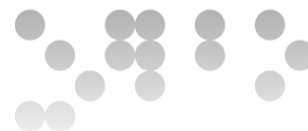
Es vol intentar millorar-la i es crea el següent índex:

```
CREATE INDEX IDX_PhoneNumber ON Person(phoneNumber);
```

En repetir la consulta s'observa que el cost d'execució continua essent el mateix. Quin creieu que pot ser el motiu, si sabem que hi ha centenars de milers de línies telefòniques i que la majoria comencen per 93? Responeu argumentadament.

SOLUCIÓ:

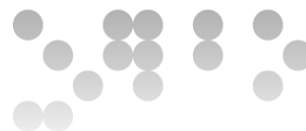
L'optimitzador calcula que el nombre de files a les que s'ha d'accedir és molt alt en relació al total de registres de la taula, per tant, el més òptim és llegir blocs sencers en cada operació d'E/S, que no fer múltiples operacions a disc per accedir per l'índex a cada registre.



EXERCICI 4

Com a responsables de seguretat informàtica d'una empresa ens plantegen de solucionar diverses problemàtiques:

- a) Es desitja que els usuaris del departament de comptabilitat puguin veure totes les dades comptables de l'empresa, a excepció de les partides de sous i salaris. Explica com ho implementaries des de la base de dades.
- b) Es desitja quedi reflectida en la taula 'auditoria' tota modificació que es faci en la taula que emmagatzema els sous dels treballadors. En concret, es desitja que s'emmagatzemi l'identificador de l'usuari que ha efectuat l'operació, el sou que existeix abans de l'actualització i el nou sou. Explica com creus que s'hauria d'implementar la funcionalitat esmentada.
- c) Es desitja millorar la qualitat de les dades que s'introdueixen en la taula 'empleats'. Quin mecanisme s'ha d'utilitzar per assegurar que els NIFs que s'introdueixen de les persones siguin correctes? En quins casos s'hauria d'executar?
- d) Supposem aquests tres casos:
 - Tenim una base de dades amb una taula 'fotos' i volem que només hi accedeixin els nostres amics.
 - Ens hem adonat que les fotos de la taula anterior les podríem classificar en públiques, reservades i confidencials. Per defecte, tots els amics que tenen accés poden veure les públiques, i hem de determinar quins poden veure també les reservades i quins les confidencials.
 - En una base de dades d'una clínica es disposa de dues taules de pacients, una amb les dades generals, que la consulten tant el personal sanitari com l'administratiu i l'altra, amb informació molt confidencial com és el cas de la història clínica, que només pot



consultar-la el personal sanitari. Quin control d'accés creus que cal aplicar en cada cas?

SOLUCIÓ:

- a) S'hauria d'utilitzar vistes per mostrar les dades. Els usuaris del departament de comptabilitat consultarien les dades mitjançant una vista que tingui l'esmentada limitació de consulta. Seria un cas típic de d'utilització de vistes com interfície per proporcionar independència entre dades i aplicacions.
- b) S'hauria de crear un disparador que s'executés (before) cada vegada que es fes una modificació del camp sou dels treballadors. Des d'aquest s'emmagatzemaria a la taula 'auditoria' el contingut de la variable d'entorn user, que conté el identificador de l'usuari, el valor del camp sou del registre que es vol modificar (OLD) i el nou valor (NEW).
- c) S'hauria de crear un disparador que donat un NIF comprovés que la lletra és correcta. El disparador hauria de ser before, i executar-se per al cas d'insercions i modificacions.
- d) En el primer cas amb un control discrecional (DAC) en tenim prou. S'ha de definir el nivell d'accés per l'usuari (o grups d'usuaris), per a tot el conjunt de dades.

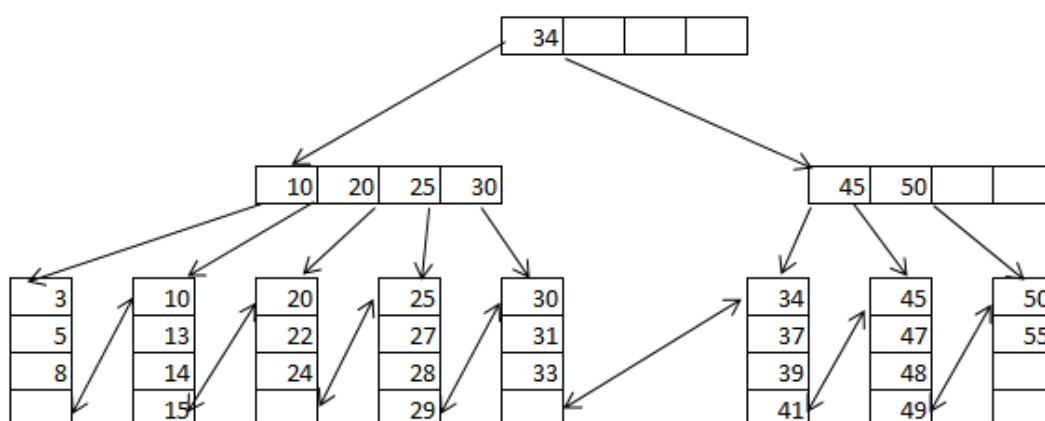
En el segon cas cal un control d'accés obligatori (MAC), ja que cada dada (foto) té assignada el seu nivell d'accessibilitat. No hem d'oblidar que el MAC requereix un DAC previ. Per tant s'aplicaran tots dos.

Per últim, el tercer cas se solucionaria amb un DAC, ja que no cal fer distincions entre usuaris per accedir a determinades dades d'una mateixa taula.



EXERCICI 5

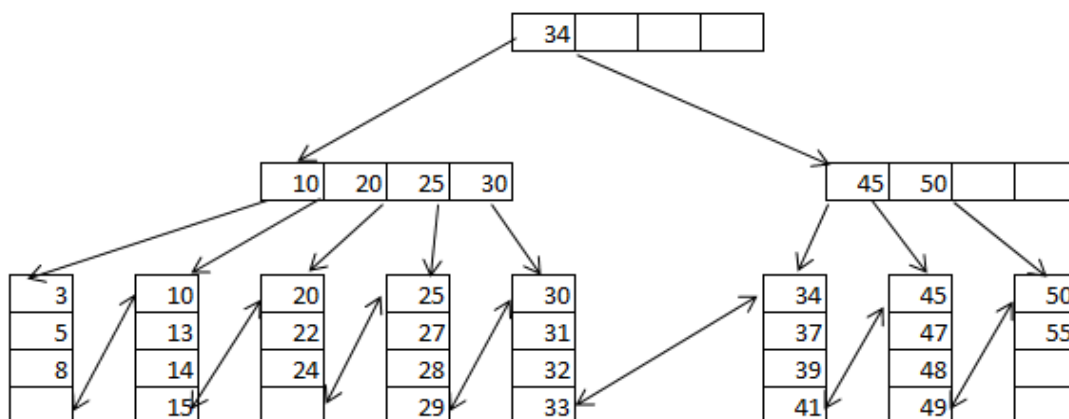
A la següent figura tenim la representació d'un arbre B+ d'ordre 2:



Mostreu com quedaria l'arbre després de fer de forma consecutiva les operacions següents. Justifiqueu breument cada resposta:

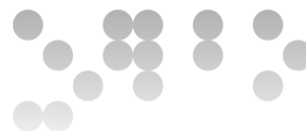
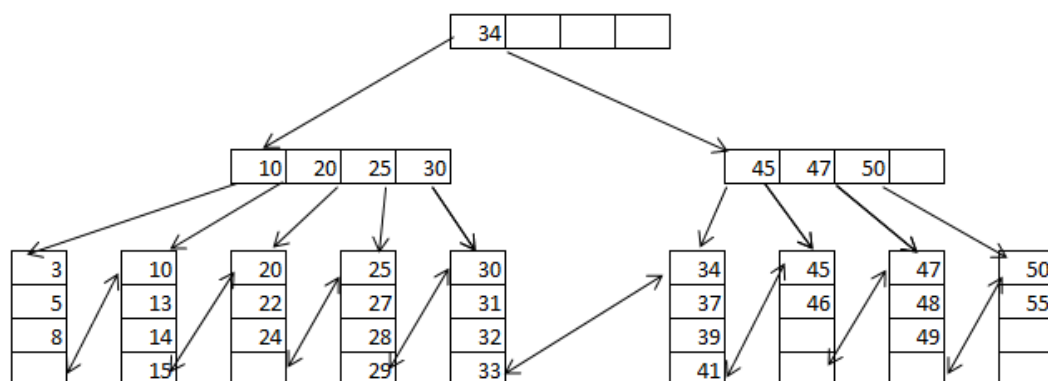
Inserir el 32:

SOLUCIÓ:



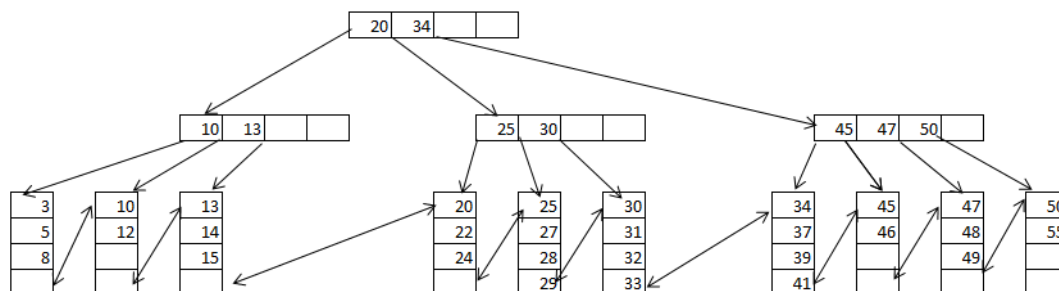
Es col·loca directament entre el 31 i el 33. Com hi cap a la fulla, no hi ha cap canvi en l'estructura i disposició dels nodes.

Inserir 46:

**SOLUCIÓ:**

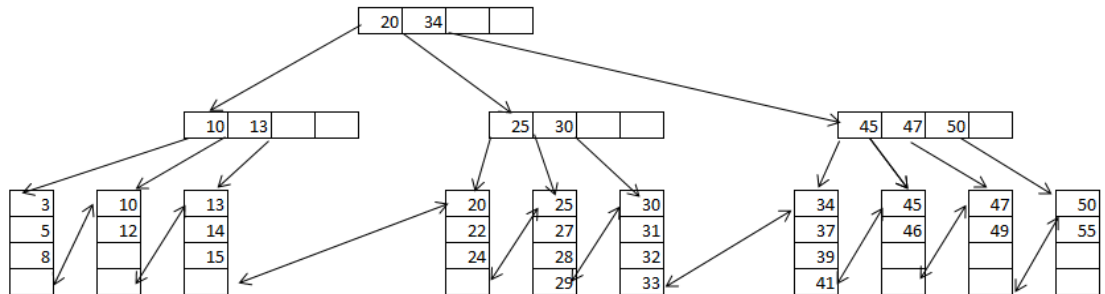
S'ha de dividir la fulla, els primers d(2) valors (45 i 46) al node de la dreta i la resta (47, 48 i 49) a l'esquerra. També cal posar un nou valor al node pare (47).

Inserir 12:

SOLUCIÓ:

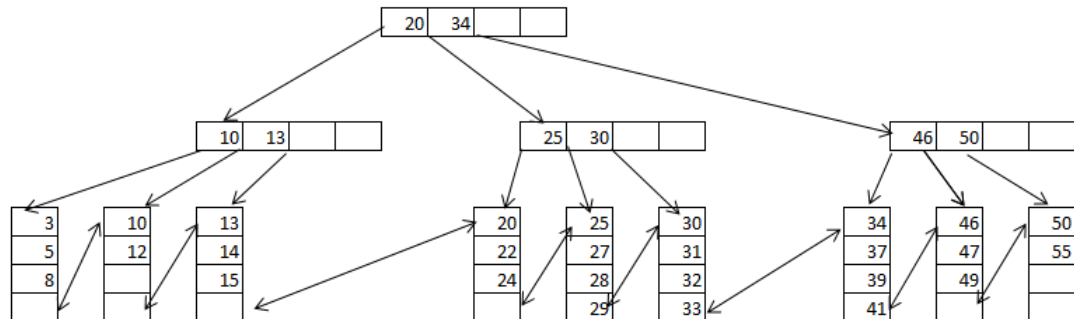
S'ha de dividir la fulla, els primer d(2) valors (10 i 12) a la dreta i la resta (13,14 i 15) a l'esquerra. També cal posar un nou valor al node pare (11). Com que el node pare és ple també es divideix (seguint el mateix criteri que al node fulla) i, per últim, s'ha de posar un nou valor a l'arrel (20).

Eliminar 48:

**SOLUCIÓ:**

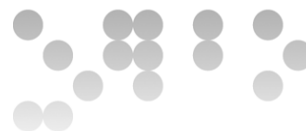
Es reorganitza el node fulla pujant un lloc el 49.

Eliminar 45:

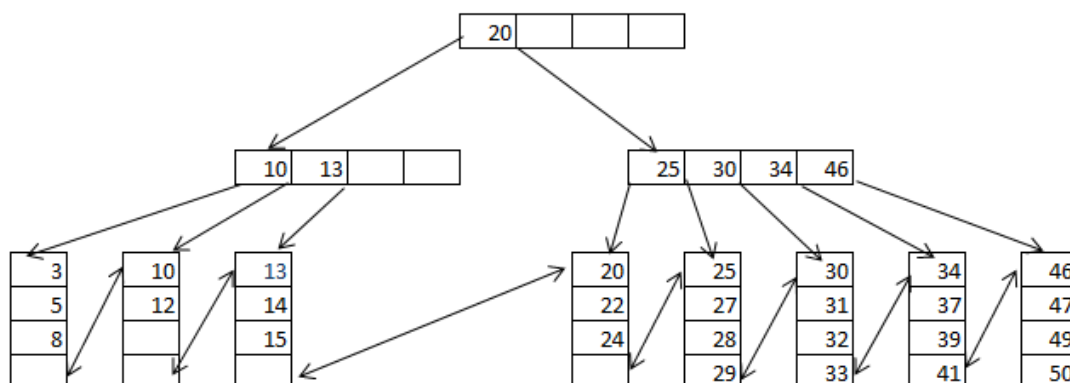
SOLUCIÓ:

El node queda només amb un valor (46), per tant s'ha de fusionar amb seu germà de la dreta. Com a conseqüència el node pare perd un valor.

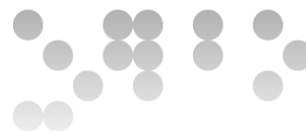
Eliminar 55:



SOLUCIÓ:



El node queda només amb un valor (50), per tant s'ha de fusionar amb seu germà de l'esquerra, ja que no té germà a la dreta. El mateix valor es perd al node pare, quedant només una posició ocupada, per tant també s'ha de fusionar amb el node pare germà de la dreta.



Recursos

Per solucionar aquesta PAC és necessari utilitzar els continguts dels mòduls 4 i 5 del material docent.

Criteris de valoració

El pes dels exercicis en la nota total de la PAC és el següent:

- Exercicis 1, 2 i 3: 15% cadascun.
- Exercici 4: 35%
- Exercici 5: 20%

Aquesta PAC té un pes del 50% en la nota d'avaluació continuada.

Format i data de lliurament

El format del fitxer ha de ser Word o OpenOffice, i s'ha de lliurar una versió del mateix fitxer en PDF. Feu enviaments independents per a la versió doc/odt i el fitxer PDF.

El nom del fitxer tindrà el format següent:

Cognom1_Cognom2_Nom.extensió

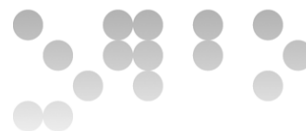
Els cognoms s'escriuran sense accents. Per exemple, un estudiant que es digui Alfredo García Melgar posaria el següent nom a l'arxiu:

Garcia_Melgar_Alfredo.doc (u .odt) i .pdf

IMPORTANT: El nom i cognoms de l'estudiant també han d'aparèixer a la portada del document amb la solució.

És responsabilitat de l'estudiant assegurar-se que els documents lliurats s'han pujat correctament i corresponen a l'activitat que cal presentar.

La data límit per lliurar la PAC 2 és dilluns dia 3 de juny .

**Nota: Propietat intel·lectual**

Sovint és inevitable, en produir una obra, fer ús de recursos creats per terceres persones. És per tant comprensible fer-ho en el marc d'una pràctica dels estudis del Grau d'Informàtica, sempre que això es documenti clarament i no suposi plagi en la pràctica.

Per tant, en presentar una pràctica que faci ús de recursos aliens, s'ha de presentar juntament amb ella un document en què es detallin tots ells, especificant el nom de cada recurs, el seu autor, el lloc on es va obtenir i el seu estatus legal: si l'obra està protegida pel copyright o s'acull a alguna altra llicència d'ús (CreativeCommons, llicència GNU, GPL ...). L'estudiant haurà d'assegurar-se que la llicència que sigui no impedeix específicament seu ús en el marc de la pràctica. En cas de no trobar la informació corresponent haurà d'assumir que l'obra està protegida pel copyright.

Hauran, a més, adjuntar els fitxers originals quan les obres utilitzades siguin digitals, i el seu codi font si correspon.

Un altre punt a considerar és que qualsevol pràctica que faci ús de recursos protegits pel copyright no podrà en cap cas publicar-se en altres mitjans, a no ser que els propietaris dels drets intel·lectuals donin la seva autorització explícita.