

# CC3201/CC42A/CC55A - Bases de Datos

## Auxiliar 7: Índices y Optimización de Consultas

Profesor : Claudio Gutiérrez  
Prof. Aux.: Violeta Chang C.

19 de Octubre de 2009

### 1. Índices

1. Qué organización utilizaría en cada uno de los siguientes casos? Fundamente su respuesta.
  - Buscar registros dentro de un intervalo
  - Buscar registros cuyos atributos posean valores específicos
2. Siempre los índices hacen más rápidas las consultas? Sea  $R(A, B, C)$  una tabla con un índice agrupado sobre A y uno desagrupado sobre B. Dar una consulta de ejemplo para cada uno de los siguientes casos:
  - Una consulta que sea más lenta al usar los índices
  - Una consulta que sea más rápida al usar los índices
  - Una consulta que no se vea afectada al usar o no los índices
3. Ordene, según el costo de acceso para una consulta del tipo  $\sigma_{atributo=constante}$ , el hecho de aplicarla a una tabla de las siguientes características: (Considerar que *atributo* tiene valores únicos y repetidos)
  - Desordenada sin índices
  - Ordenada sin índices
  - Índice hash no agrupado sobre *atributo*
  - Índice hash agrupado sobre *atributo*
  - Índice  $B^+$  sobre *atributo*

### 2. Optimización de Consultas

1. Considerando que no hay índices, optimice heurísticamente la siguiente consulta:

```
SELECT desc
FROM producto, cliente, pedido
WHERE cant>100 AND ciudad='Concepción' AND producto.idp=pedido.idp AND cliente.idc=pedido.idc
```

2. Considere el siguiente esquema de base de datos

$R_1(A, B, C, D)$  ,  $R_2(A, E, F)$  ,  $R_3(B, G)$  ,  $R_4(G, H)$

Y considere la siguiente consulta:

```
SELECT R1.B, R1.D
FROM R1, R2, R3, R4
WHERE R4.H=valorH AND R2.F=valorF AND R4.G=R3.G AND R3.B=R1.B AND R2.A=R1.A
```

Además, se entrega la siguiente información:

- Las tablas  $R_1$ ,  $R_2$ ,  $R_3$  y  $R_4$  tienen 4000, 500, 1000 y 100 registros respectivamente
- En la tabla  $R_4$  se tienen 10 valores de G para cada valor de H (en promedio). Esta proporción también se cumple para la tabla  $R_3$  (por cada 10 valores de B, hay un valor de G)
- Se sabe que un 40% de los valores de F son mayores de *valorF*
- En promedio, cada valor de B se repite 4 veces y cada valor de A se repite 10 veces en la tabla  $R_1$
- El campo A es clave en la tabla  $R_2$

Entonces, se pide que:

- Proponga un árbol de evaluación y su equivalencia en álgebra relacional
- Calcule la cantidad de celdas a leer y escribir, asumiendo que los *joins* se realizan por fuerza bruta
- Optimice la consulta en el caso de no existir índices y muestre su equivalencia en álgebra relacional
- Calcule la cantidad de celdas a leer y escribir luego de la optimización, asumiendo que los *joins* se realizan por fuerza bruta
- Desarrolle un plan de índices para esta consulta. Fundamente sus decisiones