



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
DEPARTAMENTO DE CIENCIA DE LA COMPUTACIÓN
IIC2413 - BASES DE DATOS

Taller evaluado 3 Índices

Fecha: 5 Junio de 2024

1º semestre 2024 - Profesores: N. Alvarado - A. Reyes - E. Bustos - C.Álvarez

INSTRUCCIONES

- I Al finalizar el taller debe digitalizarlo y subirlo a Canvas correctamente. No seguir esta instrucción los expone a un descuento de hasta 5 décimas.
- II Cada hoja debe contener su Nombre Completo y Número de Lista en la esquina superior izquierda.
- III Debe firmar la lista de asistencia.
- IV Duración 1 hora, CON APUNTES DE CLASE

Enunciado

Dada la tabla de datos **Clientes** conformada por 1.000 tuplas

Cliente	Email	Teléfono	Dirección	Comuna
Agustin Matte	Agustin.Matte@bbdduc.utopia	56911904638	Calle Estado 456	Santiago
Agustin Reyes	Agustin.Reyes@bbdduc.utopia	56972317149	Calle Prat 789	Iquique
...

donde ... indica que hay más datos en la tabla **Clientes**

- (1pt) Indique un posible esquema para la tabla **Clientes** (cliente, email y teléfono son atributos únicos y no nulos). R: Clientes(cliente string 30 unique not null, mail string 100 unique not null, telefono int unique not null, direccion string 100, comuna string 30)

- (1pt) Defina un surrogate key llamada **Key** para la tabla **Cientes** como llave primaria e inclúyala en el esquema. R: key serial PRIMARY KEY
- (2pt) Defina un índice para la surrogate key llamada **Key-index** y proponga una posible función de Hash para un factor de carga 1. R: CREATE INDEX key-index ON Cientes (key), $F(\text{key}) = \text{key} \bmod 100$ También es válido que digan que la PK ya está indexada.
- (2pt) Defina un índice secundario para el atributo email llamado **mail-index**. ¿Si la función de hash es $f(\text{mail}) = \text{Largo del string (mail)}$ que efecto tendría sobre el factor de carga? R: CREATE INDEX mail-index ON Cientes (mail) la función Largo(mail) es mínimo 4 y máximo 100 y puede estar muy concentrada por lo que debería podría haber mucho overflow
- (4pt) Construya tres diagramas, uno para la tabla y uno para cada uno de los dos índices con 10 registros por bucket.

NOTA: Por razones de espacio puede abreviar los registros en las tablas de hash de la siguiente forma

Registro	Registro	...	Registro
----------	----------	-----	----------

Individualizando cada registro (Por ejemplo Registro1) Key-Index

1	Ra,Ri,...,Rj
2	Rk,Rt,...,Rz
...	...
100	Ru,Rv,...,Rw

Usando un serial todos los buckets están Mail Index

1	Ra,Ri,...,Rj
2	
...	...
n	Ru,Rv,Rw

Usando cualquier función que definan probablemente los buckets

estarán con distinto número de registros y hasta overflow.

- (2pt) ¿Cómo haría Ud. para que la tabla de Hash **mail-index** no repita cada tupla completa de la tabla **Cientes**?

Se el registro puede contener el valor de la llave y un puntero al bucket del otro índice o un puntero al registro de la tabla

Nombre: _____ No de Lista: _____