



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

Taller 8 - Índices

Fecha: 28 de mayo de 2025

1^{er} semestre 2025 - Profesores: Eduardo Bustos - Christian Álvarez

Pregunta 1

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.

- I Indique un posible esquema para la tabla Clientes (cliente, email y teléfono son atributos únicos y no nulos).
- II Defina un surrogate key llamada Key para la tabla Clientes como llave primaria e inclúyala en el esquema.
- III Defina un índice para la surrogate key llamada Key-index y proponga una posible función de Hash para un factor de carga 1
- IV 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?
- V Construya tres diagramas, uno para la tabla y uno para cada uno de los dos índices con 10 registros por bucket
- VI ¿Cómo haría Ud. para que la tabla de Hash mail-index no repita cada tupla completa de la tabla Clientes?

Pregunta 2

Sea la relación $R(a, b, c, d)$ cuyo tamaño es de 1 millón de tuplas, en que cada página contiene P tuplas. Las tuplas de R están ordenadas de manera aleatoria. El atributo a es además un candidato a llave primaria, cuyos valores van del 0 al 999.999 (distribuidos uniformemente). Para cada una de las consultas a continuación, diga el número de I/O que se harán en cada uno de los siguientes casos:

- Analizar R sin ningún índice.
- Usar un $B+Tree$ unclustered sobre el atributo a . El árbol es de altura h y cada página contiene M punteros ($M > P$).
- Usar un $B+Tree$ clustered sobre el atributo a . El árbol es de altura h y cada página de hoja está ocupada al 60 %.
- Usar un $Hash Index$ unclustered con 1 millón de buckets. Cada página del índice contiene M punteros ($M > P$).
- Usar un $Hash Index$ clustered con 1 millón de buckets.

Las consultas son:

- I Encontrar todas las tuplas de R .
- II Encontrar todas las tuplas de R tal que $a < 50$.
- III Encontrar todas las tuplas de R tal que $a = 50$.
- IV Encontrar todas las tuplas de R tal que $a > 50$ y $a < 100$.