

## Guía Índices

**Problema 1)** 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 ordenados 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:

1. Encontrar todas las tuplas de  $R$ .
2. Encontrar todas las tuplas de  $R$  tal que  $a < 50$ .
3. Encontrar todas las tuplas de  $R$  tal que  $a = 50$ .
4. Encontrar todas las tuplas de  $R$  tal que  $a > 50$  y  $a < 100$ .

**Solución)** Los costos son los siguientes:

Query	Sin índice	B+Tree - u	B+Tree - c	Hash Índice - u	Hash Índice - c
$R$	$\frac{10^6}{P}$	$h + (\frac{10^6}{M} - 1) + 10^6$	$h + \frac{10^6}{0,6P} - 1$	$2 \cdot 10^6$	$10^6$
$a < 50$	$\frac{10^6}{P}$	$h + (\frac{50}{M} - 1) + 50$	$h + (\frac{50}{0,6P} - 1)$	100	50
$a = 50$	$\frac{10^6}{P}$	$h + 1$	$h$	2	1
$50 < a < 100$	$\frac{10^6}{P}$	$h + (\frac{49}{M} - 1) + 49$	$h + (\frac{49}{0,6P} - 1)$	98	48