

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 10 records. Los records 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 entradas ($M > 10$).
- Usar un *Hash Index unclustered* 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	Hash Índice
R	10^5	$h + \frac{10^6}{M} + 10^6$	$2 \cdot 10^6$
$a < 50$	10^5	$h + \frac{50}{M} + 50$	100
$a = 50$	10^5	$h + 1$	2
$50 < a < 100$	10^5	$h + \frac{49}{M} + 49$	98