

Guía Índices

Considera la relación **Personas**(**pid** INT PRIMARY KEY, **nombre** VARCHAR(100)) cuyo tamaño es de 1 millón de tuplas, en que cada página puede contener 10 tuplas. Inicialmente las tuplas de **Persona** están ordenadas de forma aleatoria, y además, están presentes todos los valores desde 1 a 1.000.000 (una forma de verlo es pensar que **pid** es un **int** auto-incremental). Ahora tienes que calcular el costo I/O para cuatro consultas que veremos más adelante, para los siguientes 5 casos:

- Analizar la tabla **Personas** sin ningún índice.
- Usar un *B+Tree clustered* sobre el atributo **pid**. puedes suponer que el árbol es de altura h . Asuma que las páginas están al 100 % de capacidad.
- Usar un *B+Tree unclustered* sobre el atributo **pid**. puedes suponer que el árbol es de altura h y que en una página caben M punteros con $M > 10$. Asuma que las páginas están al 100 % de capacidad.
- Usar un *Hash Index clustered* en donde cada tupla está en un casillero distinto.
- Usar un *Hash Index unclustered* en donde cada tupla está en un casillero distinto.

Las consultas son:

1. Encontrar todas las tuplas de la tabla **Personas**.
2. Encontrar todas las tuplas de la tabla **Personas** tal que $\text{pid} < 64$.
3. Encontrar todas las tuplas de la tabla **Personas** tal que $\text{pid} = 2048$.
4. Encontrar todas las tuplas de la tabla **Personas** tal que $100000 \leq \text{pid} < 150000$.

Solución) Los costos son los siguientes:

Query	Sin índice	B+Tree C	B+Tree U	Hash Índice C	Hash Índice U
Personas	10^5	$h + 10^5$	$h + \frac{10^6}{M} + 10^6$	10^6	$2 \cdot 10^6$
pid < 64	10^5	$h + \frac{64}{10}$	$h + \frac{64}{M} + 64$	64	$2 \cdot 64$
pid = 2048	10^5	h	$h + 1$	1	$2 \cdot 1$
$100000 \leq \text{pid} < 150000$	10^5	$h + \frac{50000}{10}$	$h + \frac{50000}{M} + 50000$	50000	$2 \cdot 50000$