

OBSERVACIONES-LAB6

Ana Sofia Villa Benavides (201923361)

Daniela Alejandra Camacho Molano (202110974)

PASO 2: Implementar modificaciones en el Reto No. 2

a) Teniendo en cuenta cada uno de los requerimientos ¿Cuántos índices implementaría en el Reto? y ¿Por qué?

NO= Número de obras, NA= número de artistas

En total usaríamos 7 índices distribuidos de la siguiente forma:

Para artistas (3):

- 1) Índice = Constituent ID: Se utiliza para la carga de datos **req 0** para hacer la referencia entre obras y artistas. Antes, esta referencia era $O(NO)*O(NA)$. Ahora es solo $O(NO)$ ya que al tener los ID como índice para encontrar el artista de un ID específico se hace mp.get que es $O(1)$ en vez de recorrer todos los artistas que es $O(NA)$
- 2) Índice= Año Nacimiento: se utiliza para el **req 1** ya que se pide listar cronológicamente los artistas, esto resulta mucho más sencillo si cada año de nacimiento o (Begin Date) es una llave que guarda todos los artistas con este Begin date. En el recorrido de buscar los artistas en el rango ya no hay que recorrer todos los artistas si no las llaves que pueden ser menos.
- 3) Índice= Nombre: se usa para el req 2 ya que ingresa el nombre y hay que buscar al artista por esta llave.

Para obras (4):

- 4) Índice= Medio: se utiliza en el **req 3** ya que pide organizar las obras de un artista por técnica.
- 5) índice = Nacionalidad: Para el **Req 4** realizamos la conexión entre obras-artistas y guardamos en una lista las nacionalidades de los artistas de cada obra. De este modo resulta útil tener las llaves de nacionalidad con las diferentes obras.
- 6) Índice= Fecha de Adquisición: se utiliza en el **req 2** ya que se pide listar cronológicamente las obras, esto resulta mucho más sencillo si cada fecha es una llave que guarda todos los artistas con este Begin date. En el recorrido de buscar los artistas en el rango ya no hay que recorrer todos los artistas si no las llaves que pueden ser menos.
- 7) Índice= departamento: se utiliza para el **Req 5**. Porque piden transportar las obras de un departamento. Al tener todas las obras de un departamento bajo una misma llave esto se facilita

b) Según los índices propuestos ¿en qué caso usaría Linear Probing o Separate Chaining en estos índices? y ¿Por qué?

De manera general, se tendría en cuenta los requerimientos y las salidas que se necesitan. La ventaja del separate chaining es que se puede acceder a toda la información recorriendo con $O(1)$ ya que se conoce la llave y con esto la operación tendría una función simple. Luego de acceder al bucket se tendrían las complejidades de una TAD lista, sin embargo, cuando se accede a la llave se tiene que hacer un recorrido para sacar información.

El Linear Probing se tiene diversas llaves y a partir de esta se avanza hasta que no se genere una colisión, esto puede generar una ventaja debido a que cuando se llega al valor dependiendo de la llave, el extraer información sería $O(1)$ debido a que es un solo elemento. Sin embargo, cuando el

número de colisiones es muy grande se puede generar un mapa muy grande y por ello ocuparía más espacio y al recorrerlo tendría una complejidad mayor,

c) Dado el número de elementos de los archivos MoMA, ¿Cuál sería el factor de carga para estos índices según su mecanismo de colisión?

Según lo visto en clase, para linear probing se aconseja utilizar un factor de carga de 0.5 y para separate chaining se utiliza 4. Por lo tanto, estos serían los factores de carga utilizados dependiendo de lo que obtengamos en los datos de tiempo.

Paso 3: Separate Chaining o Linear Probing

d) ¿Qué diferencias en el tiempo de ejecución notan al ejecutar la carga los datos al cambiar la configuración de Linear Probing a Separate Chaining?

Carga de datos	
Separate Chaining (4.00)	Linear Probing (0.50)
17.11458333	18.94270833

****como ya tenemos en la carga de datos el resto de los índices que necesitamos puede que no se vea tanto el efecto de cambiar estos dos índices.**

En general, no se evidencia un cambio grande al cargar los datos con la configuración de linear probing o separate chaining. Sin embargo, según los tiempos se muestra que Separate Chaining muestra un tiempo de carga más eficiente.

Paso 4: Modificar el factor de carga

Carga de datos			
FC	Separate Chaining	FC	Linear probing
2	15.57291667	0.2	19.47395833
8	17.24609375	0.8	19.94791667

e) ¿Qué configuración de ADT Map escogería para el índice de técnicas o medios?, especifique el mecanismo de colisión, el factor de carga y el numero inicial de elementos.

Se escogería Linear Probing debido a que en el requerimiento 3 se necesita saber que medios uso un artista en específico. Debido a que la llave tiene un único valor se puede recorrer todas las llaves y a partir de esta determinar si el artista uso esta técnica. Por otro lado, el factor de colisión se generaría cada que se repita una técnica, el factor de carga tendría que ser 0.5 ya que debe hacer un número grande de comparaciones. El número inicial de elementos sería el total de artistas.

f) ¿Qué configuración de ADT Map escogería para el índice de nacionalidades?, especifique el mecanismo de colisión, el factor de carga y el numero inicial de elementos

Se escogería Separate Chaining debido a que en el requerimiento 4 se pide clasificar a los artistas por nacionalidad y por medio de esto se tendría en cuenta todos los artistas que pertenecen a cierta nacionalidad. Las colisiones se generan cada que una nacionalidad se repita, por ello se usaría un factor de carga de 2. Por otro lado, el número inicial de elementos sería el total de nacionalidades (sin duplicados).