

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

Respuesta:

Los mapas/índices que implementaríamos en el reto son:

1. Mapa con los constituent ID de los artistas:

Para los requerimientos 3, 4 y 6 necesitamos relacionar obras con las características de sus artistas.

2. Mapa con las fechas de nacimiento

En el requerimiento 1 se necesitan las fechas de nacimientos de los artistas para listarlos cronológicamente.

3. Mapa con los medios:

Es necesario para el requerimiento 3. Se puede utilizar para encontrar rápidamente las obras de arte que cada artista tiene filtrando por medio.

4. Mapa con las nacionalidades:

Es necesario para el requerimiento 4, dado que se necesita contar todas las nacionalidades de las obras.

Observación: La lista de “artworks”, en el valor de cada key del mapa de nacionalidades, tiene como elementos el objectID de las obras, no la obra completa. Esto debido a que hicimos un ajuste en el modelo para entregar este laboratorio.

5. Mapa con los departamentos:

Para el requerimiento 5 se necesita analizar las obras que posee un departamento en específico del museo.

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

Respuesta:

En los mapas se utiliza Separate Chaining debido a que se manejan grandes cantidades de datos, y aunque cada valor unitario en la tabla pesa más que en Linear Probing dado que cada valor corresponde a una lista, el factor de carga es mayor que en Linear Probing causando que al final la complejidad espacial sea menor utilizando Chaining.

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

Respuesta:

Sabemos que para Separate Chaining la complejidad es $O(\alpha)$ donde α es el Factor de Carga, por lo que decidimos tener un factor de carga de 4 para los mapas de artistas medios, fechas de nacimiento y para nacionalidades que son las tablas más grandes y un factor de carga de 2 para departamentos que es la tabla más pequeña, lo anterior con el objetivo de optimizar la complejidad temporal y espacial. Adicionalmente se puede notar que para el factor de carga elegido aún si hay varias colisiones la probabilidad de que en una misma llave exista un número considerable de colisiones es baja y la complejidad temporal sigue siendo $O(k)$, en cambio sí se reduce mucho el factor de carga la complejidad espacial si cambia notablemente.

Paso 3: Separate Chaining o Linear Probing

Pruebas de rendimiento:

1. Pruebas de rendimiento mapa nationalities

Las pruebas de rendimiento del mapa de nationalities se realizaron en la siguiente máquina

Tamaño de LISTA artworks: 748

Tamaño de mapa artistas: 1948

Keys nacionalidades: 48

Máquina	
Procesadores	Intel(R) Core(TM) i5-8265U CPU @ 1.60GHz 1.80 GHz
Memoria RAM (GB)	8 GB
Nombre del SO	Windows 10 64-bits

Tabla 1. Especificaciones de la máquina - pruebas de rendimiento mapa nacionalidades.

1.1. Archivo small (solamente carga del map de Nacionalidades)

Factor de carga	Chaining (ms)	Probing (ms)
0,5	46.875	46.875
2,0	46.875	31.25
4,0	31.25	46.875
8,0	46.875	46.875

Tabla 2. Tiempos de ejecución carga de artworks small (solamente mapa: Nacionalidades)

1.2. Archivo large

Tamaño de LISTA artworks: 138150

Tamaño de mapa artistas: 15223

Keys nacionalidades: 118

Factor de carga	Chaining (ms)	Probing (ms)
0,5	6031.25	6281.25
2,0	6015.625	6218.75
4,0	5984.375¹	6343.75
8,0	6171.875	6734.375

Tabla 3. Tiempos de ejecución carga de artworks large (solamente mapa: Nacionalidades)

¹ Factor de carga y método de colisión seleccionado para el mapa de nacionalidades

2. Pruebas de rendimiento mapa mediums:

Las pruebas del mapa de médiums se ejecutaron en la siguiente máquina:

Máquina	
Procesadores	Intel(R) Core(TM) i7-6500U CPU @ 2.50GHz, 2601 Mhz, 2 procesadores principales, 4 procesadores lógicos
Memoria RAM (GB)	16 GB
Nombre del SO	Microsoft Windows 10 Pro. 64-bits

Tabla 4. Especificaciones de la máquina - pruebas de rendimiento mapa medios.

2.1. Archivo small (solamente carga del map de Medios)

Factor de carga	Chaining (ms)	Probing (ms)
0,5	46.875	31.25
2,0	15.625	15.625
4,0	31.25	296.875
8,0	46.875	562.5

Tabla 5. Tiempos de ejecución carga de mediums small (solamente mapa: Nacionalidades)

2.2. Archivo large

Tamaño de mapa mediums: 21251

Factor de carga	Chaining (ms)	Probing (ms)
0,5	10750.0	9250.0
2,0	11234.375	33218.75
4,0	9640.625	145734.37
8,0	9484.375	87515.625

Tabla 5. Tiempos de ejecución carga de médiums versión large (solamente mapa: Nacionalidades)

Las demoras en el archivo de medio se deben a que el mapa en cada posición guarda una lista con todas las obras de ese medio. Causando un aumento significativo en la complejidad de Probing, sobretodo para factores de carga altos en dónde se presentan muchas colisiones y se remplazan valores anteriores por falta de espacio.

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

Respuesta:

Chaining es más eficiente en tiempo. Esto se debe a que en Chaining la complejidad en tiempo es apropiadamente es $O(\alpha)$ mientras que la complejidad en Linear Probing para un mismo de factor de carga es mayor. Esto debido a que habrá colisiones con peores efectos en la complejidad. Para tener complejidades similares se tiene que disminuir el factor de carga en Linear Probing, sin embargo, para compararlos en las mismas condiciones, en esta prueba se mantuvo el mismo factor de carga independiente del mecanismo de manejo de colisiones.

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.

Respuesta:

El mecanismo de colisión es Separate Chaining, el factor de carga es 4.0 y el número inicial de elementos es 25000.

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.

Respuesta:

La configuración para el índice de nacionalidades está especificada en la siguiente tabla. Con esta configuración se observaron mejores pruebas de rendimiento y menores colisiones. La capacidad del mapa inicialmente con esta configuración es 67, al cargar el mapa queda con una capacidad de 67 y con un tamaño (size) de 118.

Mecanismo de colisión	CHAINING
Factor de carga	4.0
Número inicial de elementos	250

Tabla 6. ADT Map nacionalidades