

Paula Daza, 202111276, [p.dazad@uniandes.edu.co](mailto:p.dazad@uniandes.edu.co)

Juan David Aparicio, 202116532, [j.apariciog@uniandes.edu.co](mailto:j.apariciog@uniandes.edu.co)

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

Implementaría 7 índices. Se necesita un índice por año de nacimiento de los artistas para el requerimiento 1. Para el requerimiento 2 se necesitan 2 índices: uno por CreditLine de las obras para ver cuáles fueron adquiridas por compra (Purchase) y un índice por fecha de adquisición. Para el requerimiento 3 se implementaría un índice por medio (técnica). En el requerimiento 4 se usaría un índice por nacionalidad de los artistas. Para el requerimiento 5 se usaría un índice por departamento. Y para el bono se necesitaría un índice de obras por artista.

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

Para el índice por año de nacimiento usaríamos Linear Probing porque los distintos años de nacimiento no son excesivamente muchos. Es decir, el rango de años de nacimiento no es excesivamente grande. Así, al ser relativamente pocas llaves distintas, habrán pocas colisiones y no habrá mucha necesidad de hacer rehash. Para el CreditLine, como también hay pocos tipos, usaríamos Linear Probing. Para el índice por fecha de adquisición usaríamos Separate Chaining, ya que en este caso sí hay demasiadas posibilidades, por lo tanto pueden haber muchas colisiones y no conviene hacer rehash muy seguido con un volumen tan grande de datos, por lo cual es mejor Separate Chaining. Para los medios también usaríamos Separate Chaining, ya que también hay muchos tipos. Para las nacionalidades usaríamos Linear Probing, ya que hay un número finito de posibles nacionalidades, el cual no es muy alto. Por lo cual es mejor ahorrar memoria y facilitar la consulta con Linear Probing. Para el de los departamentos también usaríamos Linear Probing, ya que hay relativamente pocos departamentos. Y finalmente, para el índice de obras por artista usaríamos Separate Chaining, porque el número de artistas distintos es muy grande, por lo cual hacerlo con Linear Probing usaría demasiada memoria, así que es mejor usar Separate 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?

Índice por año de nacimiento: 0.4. Índice por CreditLine: 0.5. Índice por fecha de adquisición: 4. Índice por medios: 4. Índice por nacionalidades: 0.4. Índice por departamentos: 0.6. Índice por obras por artista: 4.

### **Separate Chaining 4.0**

1718.75 ms

### **Linear Probing 0.5**

1656.25 ms

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

Hubo un ligero cambio, resultando mejor el Linear Probing, debido quizás a que hay relativamente un espectro pequeño de nacionalidades. La diferencia de tiempos fue de 62.5 ms.

**Separate Chaining 2.00**

1859.375 ms

**Separate Chaining 8.00**

2031.25 ms

**Linear Probing 0.20**

1515.625 ms

**Linear Probing 0.80**

1625.0 ms

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

Usaríamos Separate Chaining, con un factor de carga de 4.0, y un número inicial de elementos correspondiente a la cantidad total de obras.

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

Escogeríamos Linear Probing, con un factor de carga de 0.4 y un número inicial de elementos correspondiente a la cantidad total de artistas (que es en el peor caso el número total de nacionalidades).