

OBSERVACIONES DE LA PRACTICA

Juan Pablo Rodríguez Briceño Cod 202022764

Nicolas Perez Teran Cod 202116903

- 1) Teniendo en cuenta cada uno de los requerimientos ¿Cuántos índices implementaría en el Reto? y ¿Por qué?
 - a) **RTA:** Se crearían 4 mapas; siendo el primero para ordenar los artistas con su ConstituentID, el resto corresponderían al Req 3 ['Medium'], 4 ['Nationality'] y 5['Department'] porque estos requieren que se tome en cuenta una sección específica de un atributo que tenga la obra o el artista, esto permitiendo que sea más rápido el proceso de filtrar según el parámetro dado.
- 2) Según los índices propuestos ¿en qué caso usaría Linear Probing o Separate Chaining en estos índices? y ¿Por qué?
 - a) **RTA:** Para el Req. 3 y 5 que requiere los buscar por ["medium"] lo mejor es Linear Probing porque requiere muchos procesos y lo mejor es que se hagan lo más rápido posible, incluso aunque se utilice mas memoria, el 4 debe utilizar Separate Chaining porque no requiere hacer muchas búsquedas y como son muchos datos, podría exceder la capacidad de la memoria RAM. y no requiere hacer mucha búsqueda .
En el de los artistas se utilizaría Linear Probing para que sea mas rápida la consulta de nombres, debido a que el archivo de artistas es menor en comparación al de obras cuando a datos se refiere.
- 3) Dado el número de elementos de los archivos MoMA, ¿Cuál sería el factor de carga para estos índices según su mecanismo de colisión?
 - a) **RTA:** En Separate Chaining, como los datos son 138.112 creemos que lo mejor es que el factor de carga sea igual a 2.0 debido a que son muchos datos y se requiere que estos no terminen en un bucket con muchos elementos.
 - b) Para el caso de Linear Probing, creemos que lo ideal es que se utilice un factor de carga de 0.5 debido a que tener el doble de la cantidad de datos es demasiado, pero al mismo tiempo debemos saber que las colisiones podrían ser un gran problema si los espacios son muy reducidos, por el mismo tema de que son muchos datos para almacenar.
- 4) ¿Qué diferencias en el tiempo de ejecución notan al ejecutar al cargar los datos al cambiar la configuración de **Linear Probing** a **Separate Chaining**?
 - a) **RTA:** El tiempo de ejecución usando Separate Chaining con un factor de carga de 4 fue de 202328.125 ms mientras que, usando Linear Probing con un factor de carga de 0.5, el tiempo de ejecución fue de 201109.375 ms. Podemos concluir que Linear Probing con 0.5 en ambos mapas hace que el proceso se ejecute notoriamente más rápido que como lo fue con Separate Chaining (Cabe destacar que se uso el csv small para esta prueba)
- 5) ¿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.

- a) **RTA:** Para técnicas o medios, utilizaría una estructura de datos rápida por lo que se necesitan constantes búsquedas, entonces Linear Probing con un numero de elementos mayor a 138112 posiblemente 276439 por ser un numero primo que es cercano al doble del numero inicial de elementos, 138219, es decir, con un factor de carga de 0.5 pues tener mas del doble seria demasiado para la memoria.
- 6) ¿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.
 - a) **RTA:** Dado que el índice de nacionalidades no requiere que se consulte repetidas veces se usaría Separate Chaining. Además, se seleccionaría un numero de elementos de 200 o mayor, pero sin una diferencia grande. Posiblemente se seleccione el 211 ya que es un numero primo cercano a 200 y es mayor. Por ultimo, se pondría un factor de carga de 2. Todo esto teniendo en cuenta que se trabaje con -small