Laboratorio 6

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

Se espera usar 5 índices en el transcurso del reto, los cuales serían:

Por departamento, para poder agrupar las obras en nuestro requerimiento 5

Por medio, para obtener que medios fueron más utilizados en el requerimiento 3

Por nacionalidad, para obtener de que país hay más autores en el requerimiento 4

Por ConstituentID, para diferenciar los autores y creación de catalogo

Por ObjectID, para diferenciar las obras y creación de catalogo

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

Esperamos usar Linear probing para los índices de medios y nacionalidad y departamento, porque, usan una cantidad relativamente pequeña de keys a comparación de otros índices como el de ObjectID, por esta razón, la complejidad temporal de carga de estos índices no será muy grande, pero, la complejidad de buscar un dato será siempre constante y su valor será de 1,5 comparaciones aproximadamente.

Por otro lado, esperamos usar Separate Chaining en los indices de ConstituentID y Object, principalmente, porque son una gran cantidad de datos por lo que la complejidad de carga de datos va a ser mejor con el Separate Chaining. Adicionalmente, el Separate Chaining se caracteriza por agrupar una gran cantidad de datos en caso de colisiones y en poca memoria.

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

Tenemos previsto añadir los archivos del MoMA, con un factor de carga de 1,2, principalmente porque tenemos que cargar todos los archivos sin que uno se quede fuera, porque toda cuenta, al igual que dejamos un margen en caso de querer añadir más keys

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?:

Dados los resultado evidenciados en el Excel realizado por nosotros, evidenciamos una diferencia minima de rango de de 100 hasta 300 milisegundos entre los dos tipos de colisiones, que para nosotros no están significativa a grandes rasgos

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.

Teniendo en cuenta nuestros resultados, la mejor opción en nuestro índice de medios, corresponde a Linear probing con un factor de carga de 0.5 debido a que fue la que mostró el menor tiempo usando el archivo large.

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.

nuestro índice de nacionalidades fue más eficiente en Separate Chaining con un factor de carga de 2 debido a que fue la que mostró el menor tiempo usando el archivo large.