Laboratorio 6

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

Para el requerimiento 1 se implementaría como índice el “Begin Date” porque nos piden organizar los artistas por medio de su fecha de nacimiento en un rango de años indicado por el usuario.

Para el requerimiento 2 se implementaría como índice el “Date” porque nos piden organizar las obras por medio de su fecha en un rango de fechas indicado por el usuario.

Para el requerimiento 3 se implementaría como índice el “Display Name” el cual tuviese como valor las obras asociadas al artista, porque nos piden clasificar los artistas por su nombre al compararlo con el nombre indicado por el usuario.

Para el requerimiento 4 se implementaría como índice el “Nationality” el cual tuviese como valor los artistas asociados a cierta nacionalidad.

Para el requerimiento 5 se implementaría como índice el “Department” porque nos piden organizar las obras por medio del departamento del museo indicado por el usuario.

Para el requerimiento 6 se implementaría como índice el “Display Name” del requerimiento 3 el cual tuviese como valor las obras asociadas al artista, porque nos piden clasificar los artistas por su cantidad de obras.

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

Para “Nationality” y “Department” utilizaríamos Linear Probing ya que tienen pocas llaves (N) y por lo tanto al aplicar la tabla de hash (M = N/a) con un factor de carga menor a 0.5 (a), la tabla (M) no ocuparía tanta memoria al manejar colisiones.

Para “Begin Date”, “Date” y “Display Name” utilizaríamos Separate Chaining ya que tienen muchas llaves (N) y por lo tanto al aplicar la tabla de hash (M = N/a) con un factor de carga mayor a 1 (a), la tabla (M) no ocuparía tanta memoria al manejar colisiones.

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

Para Linear Probing el factor de carga sería 0.5 y para Separate Chaining sería 4.

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?

Resultados 1:

Separate Chaining: (123312.5 + 125015.625) /2 = 124164.0625

Linear Probing: (125359.375 + 127171.875) /2 = 126265.625

Cruzado: (125640.625 + 123312.5 + 128531.25) /3 = 125828.12

Podemos ver que Linear Probing tiende a demorarse un poco más que Separate Chaining en este caso. Y cuando cruzamos ambos métodos de colisión dio un valor intermedio.

Resultados 2:

Chaining 2.0: 123281.25

Chaining 8.0: 126406.25

Probing 0.2: 125625.0

Probing 0.8: 123656.25

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.

Para el índice de medios el mecanismo de colisión sería Separate Chaining con un factor de carga de 4 y con un número inicial de elementos de 800

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

Para el índice de nacionalidades el mecanismo de colisión sería Linear Probing con un factor de carga de 0.5 y con un número inicial de elementos de 100