**DOCUMENTO DE ANALISIS**

1. **Teniendo en cuenta cada uno de los requerimientos ¿Cuántos índices implementaría en el Reto? y ¿Por qué?**

RTA/ De acuerdo con la información brindada para cada uno de los requerimientos, se considera que 6 índices o mapas serían suficientes para cumplir de manera óptima todos los requerimientos del reto N°2. En consecuencia, para poder explicar el porqué de esta decisión, se muestra la distribución de dichos mapas en cada uno de los requerimientos:

* Requerimiento 1: Para este requerimiento se necesita UN MAPA, el cual sería el índice de años de nacimiento de los artistas. Esto implica que cada llave sea un año y los valores sean los artistas que hayan nacido en dicho año. Los artistas de cada valor estarán organizados en un TAD lista
* Requerimiento 2: Para este requerimiento se necesitan DOS MAPAS, los cuales estarían organizados de la siguiente manera:

Primero, el índice de obras por fecha calendario. Este índice va a tener por llave cada una de las fechas y por valor un TAD lista con todas las obras que se tienen registradas en esa fecha

Segundo, el índice de artistas, que tenga por llaves cada uno de los Constituent ID, y por valor un TAD lista unitario con la información del artista. Este segundo mapa es importante para el Print de los artistas asociados

* Requerimiento 3: Para este requerimiento se necesitarán dos mapas, pero solo UN MAPA NUEVO, los cuales estarían organizados de la siguiente manera:

Primero, un mapa de los ID de los artistas con las obras correspondientes a cada uno de ellos. Esto significa que el Constituent ID será la llave mientras que las obras en un TAD lista serán los valores.

Segundo, el índice de los artistas con sus IDs que se puede reutilizar del requerimiento 2

* Requerimiento 4: Para este requerimiento se considera necesario la utilización de 2 mapas (pero solo UNO NUEVO), estos mapas serían organizados de la siguiente manera:

Primero, se necesita un nuevo mapa que relacione las nacionalidades de cada uno de los artistas. Este mapa tendría por llave cada una de las nacionalidades, mientras que los valores serían TAD lista con la información de los artistas de dicha nacionalidad

Segundo, se necesita de nuevo el mapa que relacione los Constituent ID con las obras.

* Requerimiento 5: Para este quinto requerimiento se necesita UN NUEVO MAPA, el cual relacionaría cada uno de los departamentos con las obras que lo conforma. Así las cosas, cada una de las llaves sería el departamento, mientras que los valores serían TAD lista con la información de las listas
* Requerimiento 6: Para este requerimiento se consideran que NO SE NECESITAN MAPAS NUEVOS, ya que se puede trabajar reutilizando el mapa del primer requerimiento, con las fechas de nacimiento de cada uno de las artistas relacionados en TAD lista.

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

RTA/ Se considera que en términos de tiempo, no debe haber diferencias en las búsquedas y organizaciones de los MAPAS de acuerdo con la información necesitada por cada requerimiento, por lo se podría usar cualquiera de las 2 alternativas sin generar mayores diferencias de manera teórica. Sin embargo, en aras de dar una explicación mucho más completa de la explicación dad anteriormente, a continuación, se presenta una tabla mostrada en clase con los índices de crecimiento de cada una de las opciones de los mapas.:



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

RTA/ De acuerdo con la respuesta dada para el enunciado anterior, y tomando en consideración la cantidad de datos de cada uno de los archivos de MoMA, se presenta la siguiente tabla con los factores de carga para cada uno de los casos:

|  |  |  |
| --- | --- | --- |
|  | Separate Chaining | Linear Probing |
| Factor de Carga | 4.0 | 0.5 |

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

RTA/ Para poder dar una respuesta completa a esta pregunta, lo primero que se hace pertinente es la especificación del ambiente de pruebas, como se muestra a continuación:

|  |  |  |
| --- | --- | --- |
|  | Máquina 1 | Máquina 2 |
| Procesadores | Intel(R) Core(TM) i7-7500U CPU @ 2.70GHz 2.90 GHz | AMD RYZEN 5700U 1.80GHz |
| Memoria RAM (GB) | 8,00 GB | 16,00 GB |
| Sistema Operativo | Windows 10 –x64 bits | Windows 10 x64bits |

Posteriormente, se muestran las pruebas que se realizaron en cada uno de los computadores para este enunciado:

Maquina 1:

|  |  |  |  |
| --- | --- | --- | --- |
| **Porcentaje de la muestra [pct]** | **Tamaño de la muestra** | **Separate Chaining [ms]** | **Linear Probing[ms]** |
| small | Obras:768  Artistas:1948 | 312.5 | 171.875 |
| 10.00% | Obras:15008  Artistas:6656 | 68906.25 | 63593.75 |
| 20.00% | Obras:29489  Artistas:8724 | 308234.375 | 269312.5 |
| 30.00% | Obras:43704  Artistas:10063 | 650859.375 | 602953.125 |

Maquina 2:

|  |  |  |  |
| --- | --- | --- | --- |
| **Porcentaje de la muestra [pct]** | **Tamaño de la muestra** | **Separate Chaining [ms]** | **Linear Probing[ms]** |
| small | Obras:768  Artistas:1948 | 234.375 | 218.75 |
| 10.00% | Obras:15008  Artistas:6656 | 52671.875 | 51421.125 |
| 20.00% | Obras:29489  Artistas:8724 | 219406.25 | 211921,875 |
| 30.00% | Obras:43704  Artistas:10063 | 499812.5 | 484406,25 |

1. **¿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.**

RTA/ De acuerdo con las pruebas hechas en las 2 máquinas la configuración de ADT MAP para el caso de los medios sería la siguiente:

Mecanismo de Colisión: Linear Probing

Factor de Carga: 0.5

Número inicial de elementos: nextPrime(# de obras (dependiendo el archivo que se esté utilizando) // 3 )

1. **¿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.**

RTA/ Nuevamente, de acuerdo con las pruebas que se han hecho en cada una de las 2 máquinas, la mejor configuración sería la siguiente:

Mecanismo de Colisión: Linear Probing

Factor de Carga: 0.5

Número inicial de elementos: nextPrime(230(# aprox de países en el mundo) )

1. **Anotaciones de las pruebas con diferentes factores de carga de cada uno de los mecanismos de tratamiento de colisiones para los mapas. Las siguientes pruebas se realizaron con el 10% de los datos en el equipo reconocido anteriormente como Máquina 1:**

Separate Chaining:

|  |  |
| --- | --- |
| Factor de Carga | Tiempo(ms) |
| 2.0 | 72187.5 |
| 4.0 | 68906.25 |
| 8.0 | 72593.75 |

Linear Probing:

|  |  |
| --- | --- |
| Factor de Carga | Tiempo(ms) |
| 0.2 | 67234,375 |
| 0.5 | 63593.75 |
| 0.8 | 69015,625 |