

Reto 2 Labortario 6

-Requerimiento. 3 -- Samuel Josué Freire Tarazona, 202111460, s.freire@uniandes.edu.co

- Requerimiento. 4 -- José David Martínez Oliveros, 202116677, jd.martinezo1@uniandes.edu.co

Pregunta de análisis:

- **Teniendo en cuenta cada uno de los requerimientos ¿Cuántos índices implementaría en el Reto? y ¿Por qué?**
 - Para el primer requerimiento se crearía solo 1 índice, el cual sería de Años. Este índice, estaría compuesto de llaves, las cuales serían los años de nacimiento de los artistas, y de los valores asociados, los cuales serían los artistas que nacieron en ese año si es que son iguales.
 - Para el segundo requerimiento se crearían 1 solo índice, el cual sería de fecha de adquisición. Este índice, estaría compuesto de llaves, las cuales serían los años de adquisición, y de los valores asociados, los cuales serían las obras que fueron adquiridas en ese año.
 - Para el tercer requerimiento se crearían 2 índices. El primero sería de obras, organizadas por sus códigos. El segundo sería de artistas organizado por sus códigos. Este índice, estaría compuesto de llaves, las cuales serían los años de adquisición, y de los valores asociados, los cuales serían las obras que fueron adquiridas en ese año.
 - Para el cuarto requerimiento se crearían 1 índices. El primero sería de autores categorizados por sus nacionalidades. Este índice, estaría compuesto de llaves, las cuales serían las nacionalidades, y de los valores asociados, los cuales serían los autores nacidos en esa nación.
 - Para el quinto requerimiento se crearían 1 índices. Esta organización se haría a partir del departamento de las obras Este índice, estaría compuesto de llaves, las cuales serían los departamentos, y de los valores asociados, los cuales serían las obras de ese departamento.

- **Según los índices propuestos ¿en qué caso usaría Linear Probing o Separate Chaining en estos índices? y ¿Por qué?**
 - Para el primer índice de los años de los artistas se usaría linear probing. Porque al ser tan variados los años, es muy probable que se vaya a crear espacio innecesario si usa el separate chaining.
 - Para el segundo índice creado para los años de las obras, pasaría algo similar algo anterior. Ya que, las obras al tener fechas de adquisición tan variadas, es mejor crear campos únicos para cada fecha. Que crear varios campos grandes, donde se van a guardar solo una fecha.
 - Para el tercer índice, creado para organizar las obras por autores. Se usaría un chaining, ya que puede organizarse mejor la información porque se reducen a los autores de la obras.
 - Para el cuarto requerimiento, creado para organizar los artistas por sus códigos. Se usaría un linear probing, ya que al no tener colisiones, es mejor que cada uno se ubique en su lugar.
 - Para el quinto índice creado para las nacionalidades, se usaría un chaining. Esta sería la mejor opción porque se agruparían los autores en sus respectivas nacionalidades y permite una mejor búsqueda. Además, al haber tantas colisiones por este factor, se vuelve más manejable de esta manera.
 - Para el sexto índice, creado para los departamentos, se usaría similar a la anterior. YA que al existir tantas colisiones por este tipo de elementos. Se puede organizar de mejor manera las obras para hacer su búsqueda e investigación más fácil.

- **Dado el número de elementos de los archivos MoMA, ¿Cuál sería el factor de carga para estos índices según su mecanismo de colisión?**
 - Para el primer índice, el factor de carga sería 0.02 porque el número de llaves esperadas entre los valores totales del MoMA maso menos se podría hacer de 0.02
 - Para el segundo índice, el factor de carga sería 0.04 porque el número de llaves esperadas entre los valores totales del MoMA maso menos se podría hacer de 0.04

- Para el Tercer índice, el facto de carga seria 0.04 porque el número de llaves esperadas entre los valores totales del Moma maso menos se podría hacer de 0.04
- Para el Cuarto índice, el facto de carga seria 1.0 porque el número de llaves esperadas entre los valores totales del Moma maso menos se podría hacer de 1.0
- Para el Quinto índice, el facto de carga seria 0.1 porque el número de llaves esperadas entre los valores totales del Moma maso menos se podría hacer de 0.1
- Para el Sexto índice, el facto de carga seria 1.0 porque esperadas entre los valores totales del Moma maso menos se podría hacer de 1.0

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

- el número de llaves

	Tiempo
Chaining (4,00)	16968,75
Probing (0,5)	16718,65

- Las diferencias entre estas dos cargas es mínima. Cabe aclarar que se están usando los archivos large y se están cargando más índices aparte de estos dos. Sin embargo, se puede ver que linear probing tiende a tener un mejor rendimiento, ya que mostro un mejor número. Por lo que se podría decir que para esta carga linear probing es mejor.

- **Tomas de Tiempo para los dos índices**

	Tiempo
Chaining (4,00)	16968,75
Probing (0,5)	16718,65
Chaining (2,00)	15937,5
Chaining (8,00)	16421,88
Probing (0,2)	16125
Probing (0,8)	8609,375

- **¿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.**

- Personalmente usaría para las técnicas usaría, el ADT map de Separate Chaining, ya que me parece que agrupar todas las obras en algo que se repite mucho es lo mejor. En este caso agruparlas por técnicas, no va a tomar tanto espacio como si se pusieron por separado. Viendo los datos de la toma anterior, usuario el factor

de carga de 1.0. Y los valores iniciales serian maso menos 21000, ya que los valores de las obras o sus técnicas van maso meno por ese lado.

- **¿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.**
 - Personalmente usaría para las nacionalidad también el chaining, Por la misma a razón anterior, ay que hay varias obras que están relacionadas por sus autores con la nacionalidad. En este caso es mejor agruparlas por subconjuntos. Viendo la toma de datos, usaría el separate chaining con un factor de 1.0, ay com tamañoo esperado maso menos seria del mismo tamaño de los artistas de 20000