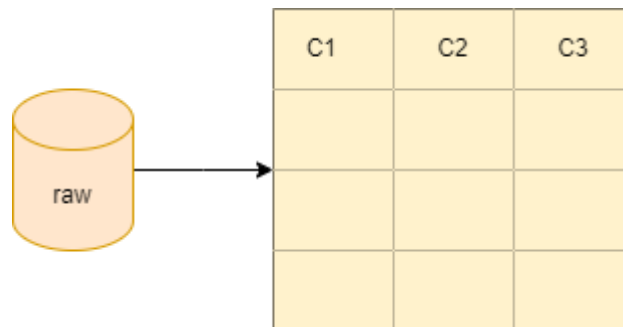


# Clase del martes 10 de octubre del 2023

## índices en bases de datos

¿Qué es un índice?

Cuando se desarrolla una base de datos se guarda dentro del disco el raw data, el raw data pasa por un tipo de transformación, normalmente una normalización, una limpieza de datos, (limpiar datos es quitar lo que no aporta a lo que estamos desarrollando), una normalización de datos es cuando se pasan todos los datos al mismo formato. A un formato consistente con lo que hay en la base de datos.



Aumentación de datos: agregar nuevos datos o sacar nuevos datos a partir de los que ya se tienen.

Aumentar, limpiar y otro son transformaciones de datos.

En una base de datos relacional tenemos los datos en tablas, en NoSQL documental se guarda en documentos.

En la base SQL se toma la información raw y se mete dentro de una o más tablas.

No es obligatorio tener normalizada una base de datos SQL porque por un lado se tiene el rendimiento y en el otro la estructuración de la base.

Una búsqueda lineal o secuencial es la más sencilla que puede realizarse, se revisa secuencialmente hasta el final y en el peor de los casos realiza un UDN.

Esta búsqueda secuencial se conoce como Scanning.

En una base de datos documental lo que se tiene es un Json. Al tratar de recuperar información de una de estas bases, se pasa de información binaria a información textual, pero no tengo seguridad de que las sean iguales entonces requiere más logísticas.

En una base de datos documental y una SQL, si solo se almacenan documentos, una SQL sería más rápida si no se utilizan índices.

cluster index: los datos están organizados en registros y la única forma de encontrar un registro es ir uno a uno revisando si es lo que busco, esto es un scanner, no es malo dependiendo de las circunstancias.

Una organización de datos es hacer algún tipo de estructura de datos que me permita hacer una búsqueda.

Un archivo de índice tiene un header en dicho archivo, que contiene metadatos acerca de los datos que estoy almacenando.

metadato más importante es el Head.

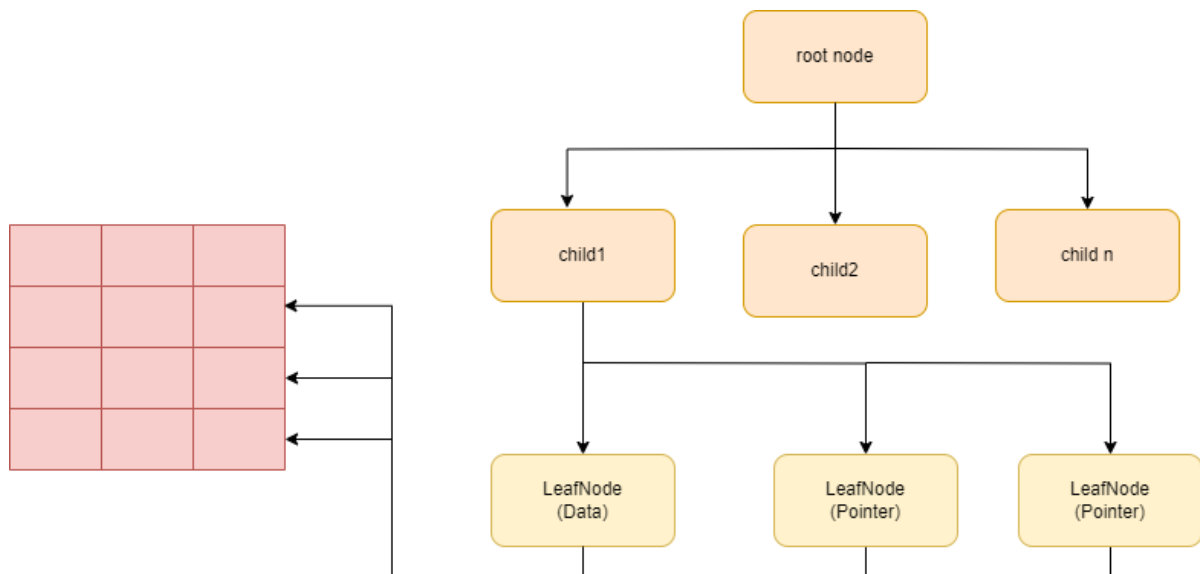
Head = -1 significa que no hay nada en la base de datos? -> revisar en internet

Índice non-cluster: crea un archivo afuera del archivo de datos que apunta hacia el archivo de datos.

Si el dataset es muy grande, es imposible guardar todo en la memoria

solo se puede poner un índice cluster por base de datos.

si tengo un dataset de 1 giga no puedo meterlo todo en memoria , hay que descender por niveles lo que implica ir a disco



Un índice non cluster es pequeño y va todo a memoria.

Un índice cluster es imposible meterlo todo si es muy grande, hay que estar yendo a disco y tarda un poco más

Una tabla de hash genera datos a partir de una función.

