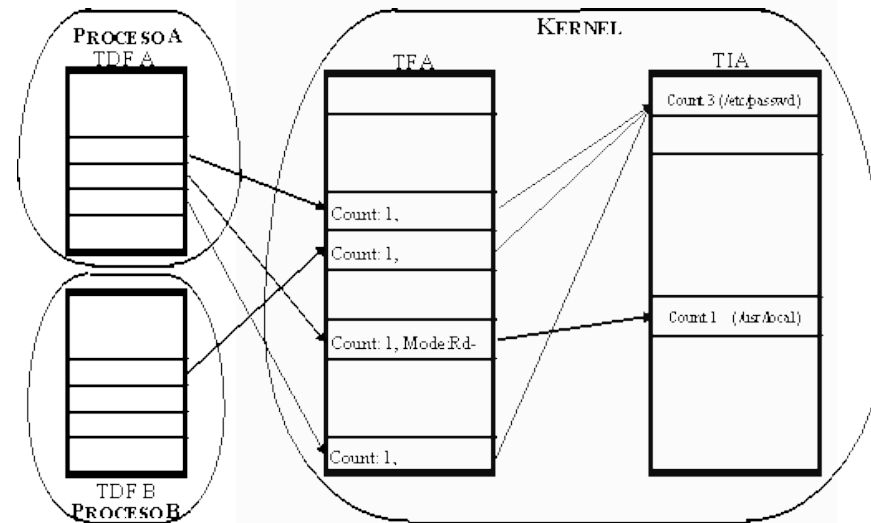


Manejo e Implementación de Archivos

Guatemala 11 de noviembre de 2021

Ing. David Luna



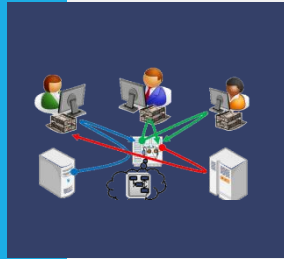
Universidad
Rafael Landívar
Tradición Jesuita en Guatemala

Agenda



Índices

Índices



Índices

¿Qué es?

- ☐ Estructura de datos que optimiza el acceso a los registros del archivo maestro.
- ☐ Debido a que almacena ciertos datos, suele caber en memoria principal y por ende su acceso es más rápido que acceder al archivo maestro.
- ☐ Debe almacenar, como mínimo, un campo para identificar al dato y la dirección física para localizarlo.
- ☐ Es responsabilidad del diseñador del sistema de archivos, definir de manera óptima los índices, su organización y su contenido.

Índices

Características generales:

- Los índices siempre deben estar clasificados (ordenados), según su atributo llave.
- Un archivo de datos puede tener asociado más de un índice.
- Un índice puede estar asociado a uno o más campos de un registro.

Índices

Características generales:

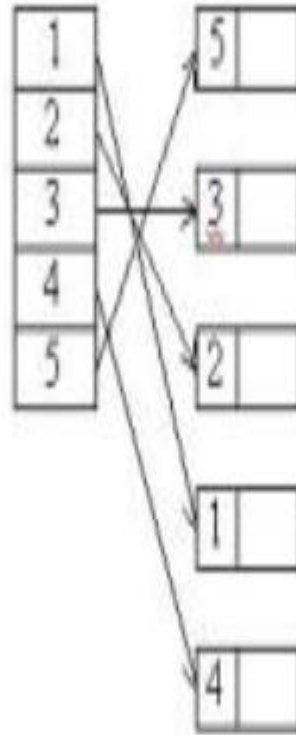
- Unicidad: es la propiedad que exige que la llave de búsqueda sea única.
- Duplicidad: es la propiedad que permite tener valores repetidos para la llave de búsqueda.



Clasificación

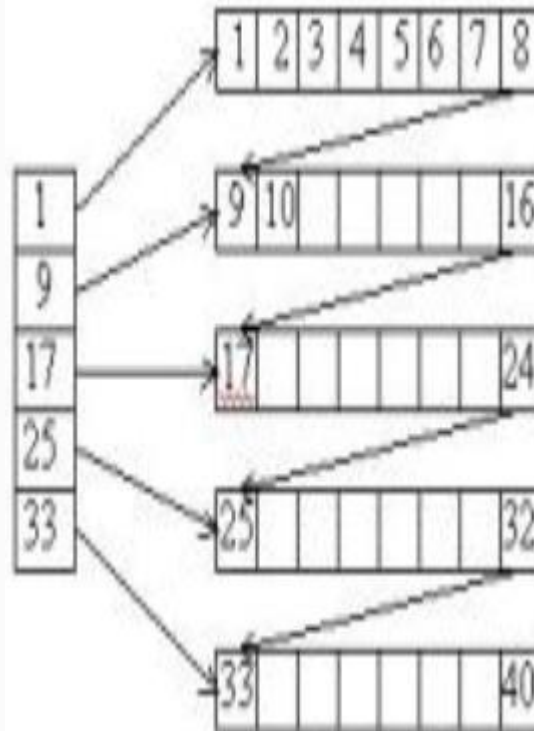
Clasificación por densidad:

● Un índice **denso** o **exhaustivo** es el que asocia una entrada en el índice, por cada registro de datos.



Clasificación por densidad:

● Un índice **disperso**, **escaso** o **selectivo** es el que no necesariamente asocia una entrada en el índice, por cada registro de datos. Pueden haber registros en el archivo de datos, que no tengan una correspondiente llave en el índice.



Clasificación por frecuencia de actualización:

- Un índice en línea es el que se actualiza en sincronización con cada operación en el archivo de datos.
- Un índice en batch se usa cuando se almacena una bitácora de transacciones y en un momento determinado se lee dicha bitácora para actualizar el índice con todas las operaciones, es decir que la actualización del archivo de datos y el índice no se hace al mismo tiempo.

Clasificación por tipo de llave:

- Un índice simple o primario

Es aquel que sólo registra la llave primaria, la cual identifica de forma única a cada registro.

El índice primario no permite la búsqueda de datos por campos que no sean la llave primaria.

Clasificación por tipo de llave:

- Un índice secundario

Incluye más campos que la llave primaria y permite llaves duplicadas, lo cual es ventajoso para las búsquedas, pero puede reducir el desempeño durante las otras operaciones

-

Clasificación por tipo de llave

Este es un archivo de datos, no tiene índice.

Código	Título	Intérprete	Año	Duración
ACalMil01	Mil horas	Andrés Calamaro	2000	03:59
MetaMot01	Motorbreath	Metallica	1983	03:09
EagHot01	Hotel California	Eagles	1972	06:31
SysCho01	Chop Suey	System Of A Down	2001	03:26
Cullle01	Illegal	Cultura Profética	2011	04:23
ACalDul02	Dulce Condena	Andrés Calamaro	2001	04:41
GonFel01	Felicidad	Gondwana	2002	04:45
MetaCre02	Creeping death	Metallica	1984	05:01
CafIng01	Ingrata	Café Tacvba	1994	03:34
ACalEst03	Estadio Azteca	Andrés Calamaro	2005	03:34
MarDon01	Donde Estará mi Primavera	Marco Antonio Solís	2008	04:13
MetaSee03	Seek and destroy	Metallica	1983	06:56
ACalMiE04	Mi enfermedad	Andrés Calamaro	2005	04:08
MrbEls01	Callaita	Bad Bunny	2019	03:15

- ¿Cuál campo podría ser la llave de un índice?
- ¿Sería posible incluir al intérprete en alguna llave?

Clasificación por tipo de llave

Índice simple o primario es aquel que guarda solamente la llave primaria, identificando de forma única a un registro almacenado en un archivo.

Índice Primario

Código	Dirección Física
ACalMil01	1
ACalDul02	6
ACalEst03	10
ACalMiE04	13
CafIng01	9
CulIle01	5
EagHot01	3
GonFel01	7
MarDon01	11
MetaCre02	8
MetaMot01	2
MetaSee03	12
MrbEls01	14
SysCho01	4

Archivo maestro

	Código	Título	Intérprete	Año	Duración
1	ACalMil01	Mil horas	Andrés Calamaro	2000	03:59
2	MetaMot01	Motorbreath	Metallica	1983	03:09
3	EagHot01	Hotel California	Eagles	1972	06:31
4	SysCho01	Chop Suey	System Of A Down	2001	03:26
5	CulIle01	Illegal	Cultura Profética	2011	04:23
6	ACalDul02	Dulce Condena	Andrés Calamaro	2001	04:41
7	GonFel01	Felicidad	Gondwana	2002	04:45
8	MetaCre02	Creeping death	Metallica	1984	05:01
9	CafIng01	Ingrata	Café Tacvba	1994	03:34
10	ACalEst03	Estadio Azteca	Andrés Calamaro	2005	03:34
11	MarDon01	Donde Estará mi Primavera	Marco Antonio Solís	2008	04:13
12	MetaSee03	Seek and destroy	Metallica	1983	06:56
13	ACalMiE04	Mi enfermedad	Andrés Calamaro	2005	04:08
14	MrbEls01	Callaita	Bad Bunny	2019	03:15

Clasificación por tipo de llave

Un índice secundario incluye más campos que la llave primaria y permite llaves duplicadas.

Índice Secundario

Llave Secundaria	Código (llave primaria)
Andrés Calamaro	ACalDul02
Andrés Calamaro	ACalEst03
Andrés Calamaro	ACaLMiE04
Andrés Calamaro	ACaLMi01
Café Tacvba	CafIng01
Cultura Profética	CulIle01
Eagles	EagHot01
Gondwana	GonFel01
Marco Antonio Solís	MarDon01
Metallica	MetaCre02
Metallica	MetaMot01
Metallica	MetaSee03
Mr Black	MrbEls01
System Of A Down	SysCho01

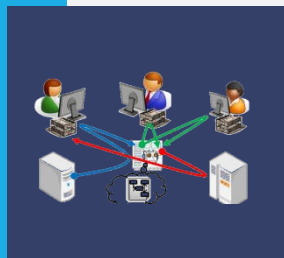
Índice Primario

Código	Dirección Física
ACaLMi01	1
ACalDul02	6
ACalEst03	10
ACaLMiE04	13
CafIng01	9
CulIle01	5
EagHot01	3
GonFel01	7
MarDon01	11
MetaCre02	8
MetaMot01	2
MetaSee03	12
MrbEls01	14
SysCho01	4

Archivo maestro

	Código	Título	Intérprete	Año	Duración
1	ACaLMi01	Mil horas	Andrés Calamaro	2000	03:59
2	MetaMot01	Motorbreath	Metallica	1983	03:09
3	EagHot01	Hotel California	Eagles	1972	06:31
4	SysCho01	Chop Suey	System Of A Down	2001	03:26
5	CulIle01	Illegal	Cultura Profética	2011	04:23
6	ACalDul02	Dulce Condena	Andrés Calamaro	2001	04:41
7	GonFel01	Felicidad	Gondwana	2002	04:45
8	MetaCre02	Creeping death	Metallica	1984	05:01
9	CafIng01	Ingrata	Café Tacvba	1994	03:34
10	ACalEst03	Estadio Azteca	Andrés Calamaro	2005	03:34
11	MarDon01	Donde Estará mi Primavera	Marco Antonio Solís	2008	04:13
12	MetaSee03	Seek and destroy	Metallica	1983	06:56
13	ACaLMiE04	Mi enfermedad	Andrés Calamaro	2005	04:08
14	MrbEls01	Callaita	Bad Bunny	2019	03:15

Archivos Indizados



Archivos Indizados

Clasificación por Estructura

- Invertido
- Binario
- Árbol B, Árbol B+, Árbol B*
- Multinivel
- Multillave
- Multianillo
- Híbrido
- ...otros...

Archivos Invertidos

- Un archivo invertido corresponde al archivo maestro que ha sido ordenado físicamente por otro atributo, diferente al atributo llave.
- El término totalmente invertido se aplica a un archivo en el que todos los atributos tienen índices. Sólo existe una llave primaria y el resto son llaves secundarias.



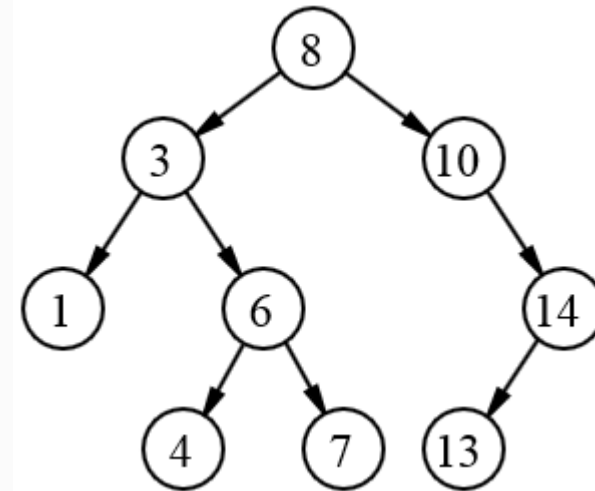
Archivo Indizado con estructura de árbol

Árbol binario de búsqueda.

Archivo Indizado con estructura de árbol

Árbol binario de búsqueda

- La mayoría de los árboles binarios son de búsqueda
- El subárbol izquierdo de cualquier nodo (si no está vacío) contiene valores menores que el que contiene dicho nodo.
- el subárbol derecho (si no está vacío) contiene valores mayores.



Archivo Indizado con estructura de árbol

Árbol binario de búsqueda

Tamaño:

9

Profundidad:

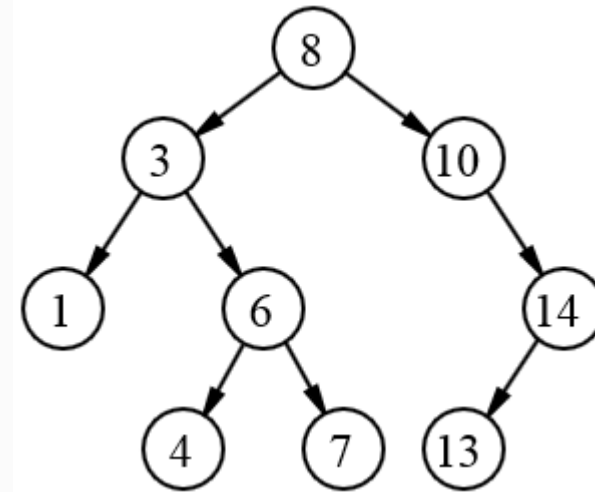
3

Raíz:

8

Hojas:

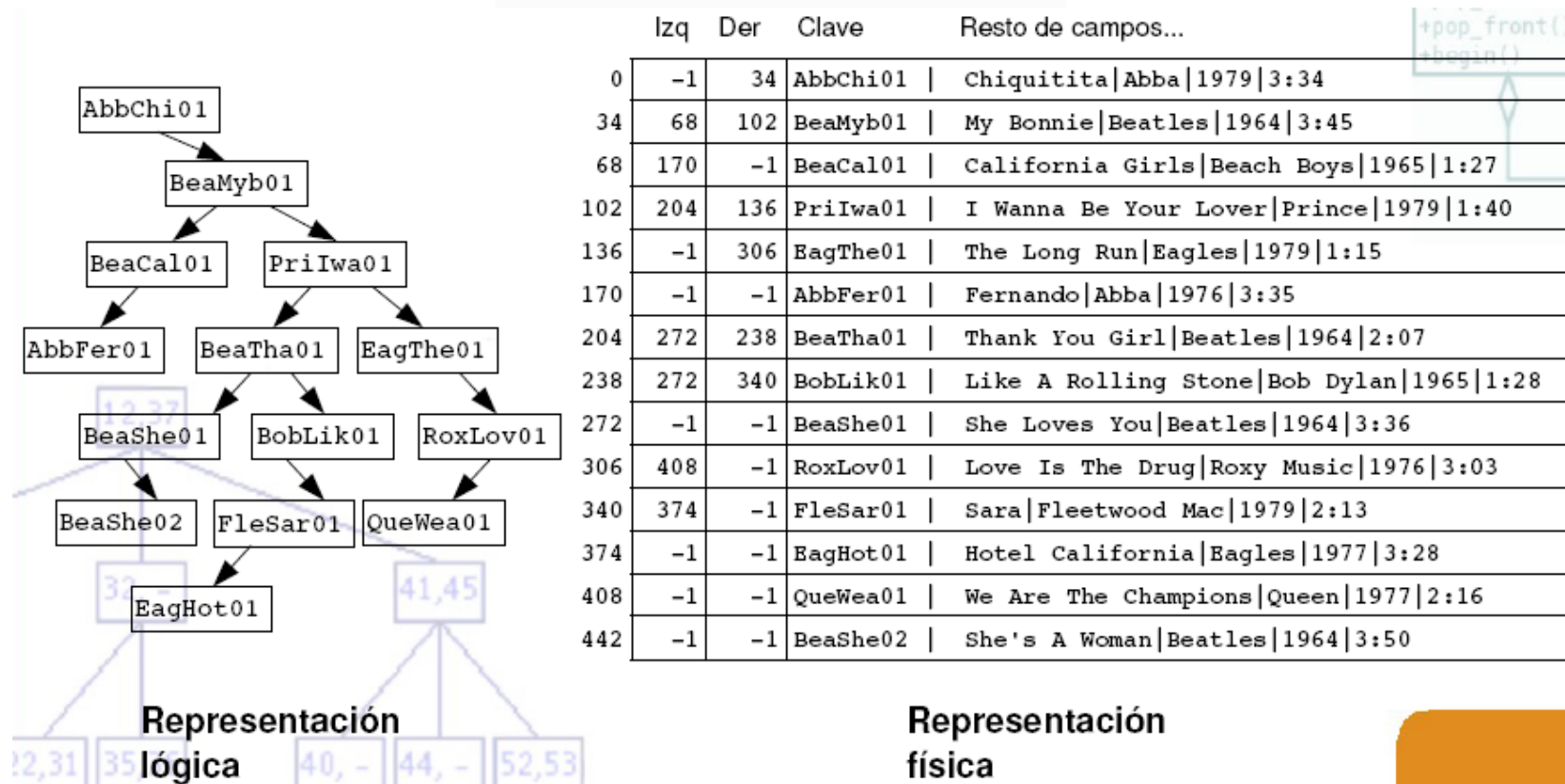
1, 4, 7 y 13



Un árbol binario de búsqueda de tamaño 9 y profundidad 3, con raíz 8 y hojas 1, 4, 7 y 13

Archivo Indizado con estructura de árbol

Ordenamiento lógico de los datos, siguiendo una estructura de árbol binario de búsqueda.



Archivos de árboles binarios

Ventajas

- Una pequeña cabecera en el archivo indica la posición del nodo raíz del árbol.
- Los nodos hoja son aquellos cuyos índices izquierda y derecha apuntan a -1.
- Representación muy compacta, pues tanto los datos como los índices están almacenados en un único archivo.
- No es necesario ordenar físicamente los datos. Los datos se insertan al final y se enganchan al nodo hoja correspondiente.

Archivos de árboles binarios

Desventajas

- Desequilibrado del árbol tras varias inserciones y borrados.
- La búsqueda binaria no es suficientemente eficiente, se requieren muchos accesos para encontrar un dato.

Ejemplo:

Se tiene el siguiente Archivo con estructura de árbol binario en donde se almacenarán los datos de los estudiantes de MEIA:

Izq.	Der.	Carnet	Nombre

Se insertan de acuerdo al listado del portal.

Ejemplo:

Insertamos carnet 2120316 :

Rep. Lógica

2120316

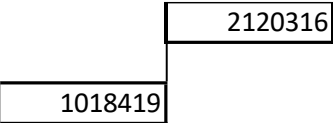
Rep. Física

Iz	Der	Carnet	Nombre
-1	-1	2120316	ALVARADO ERICK DAVID

Ejemplo:

Insertamos carnet 1018419:

Rep. Lógica



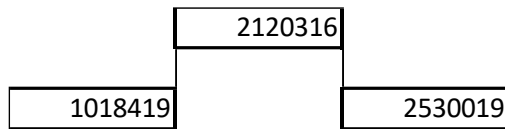
Rep. Física

Iz	Der	Carnet	Nombre
2	-1	2120316	ALVARADO ERICK DAVID
-1	-1	1018419	ALVAREZ MAZARIEGOS RAFAEL ANDRÉS

Ejemplo:

Insertamos carnet 2530019:

Rep. Lógica



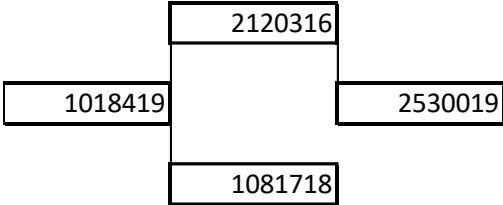
Rep. Física

Iz	Der	Carnet	Nombre
2	3	2120316	ALVARADO ERICK DAVID
-1	-1	1018419	ALVAREZ MAZARIEGOS RAFAEL ANDRÉS
-1	-1	2530019	ARAGÓN LÓPEZ DIANA ALEJANDRA

Ejemplo:

Insertamos carnet 1081718:

Rep. Lógica

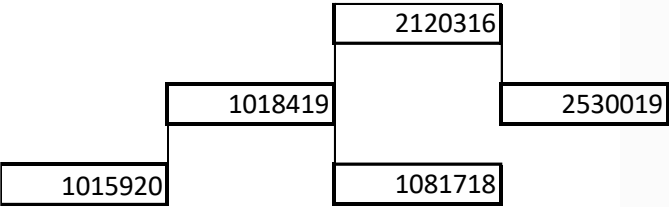


Rep. Física

lz	Der	Carnet	Nombre
2	3	2120316	ALVARADO ERICK DAVID
-1	4	1018419	ALVAREZ MAZARIEGOS RAFAEL ANDRÉS
-1	-1	2530019	ARAGÓN LÓPEZ DIANA ALEJANDRA
-1	-1	1081718	AZURDIA IZEPPi CRISTIAN JACDONI

Ejemplo:

Insertamos carnet 1015920 :
Rep. Lógica



Rep. Física

Iz	Der	Carnet	Nombre
2	3	2120316	ALVARADO ERICK DAVID
5	4	1018419	ALVAREZ MAZARIEGOS RAFAEL ANDRÉS
-1	-1	2530019	ARAGÓN LÓPEZ DIANA ALEJANDRA
-1	-1	1081718	AZURDIA IZEPPi CRISTIAN JACDONI
-1	-1	1015920	BECH FURLÁN KARL MARTIN



Ejemplo:

¿Cómo quedaría nuestro archivo con la inserción de los siguientes estudiantes?



Proyecto

Última fase...

Gracias

¿ALGUNA PREGUNTA?