

Dispersión  
extensible de  
archivos

# CONCEPTOS DE BASES DE DATOS

# DISPERSIÓN EXTENSIBLE

La **dispersión extensible** es una **técnica para manejar tablas hash cuando empiezan a crecer**.

Se usa especialmente cuando **no sabés de antemano cuántos datos vas a tener**.

En lugar de definir un tamaño fijo para la tabla y después tener que reorganizar todo (lo que sería lento y costoso), la dispersión extensible **permite que la estructura se adapte de manera flexible** al tamaño real de los datos.

Se usa un **archivo para implementar la tabla de direcciones**

Se utiliza otro archivo para almacenar cada **cubeta** que contiene varios registros o claves.

Cuando una cubeta **se llena, no se agranda la cubeta**. En cambio:

- Se **divide** esa cubeta en dos.
- Si hace falta se duplican las direcciones.

**El hash** que se calcula para cada dato **se va a ingresar**

# DISPERSIÓN EXTENSIBLE - ESTADO INICIAL

Estado inicial de un archivo con bloques con capacidad para 3 datos

Tabla de dispersión

Bits de dispersión 0

0(0): 0

Archivo de datos

#Bloque	Bits	Clave R1	Clave R2	Clave R3
0				

# DISPERSIÓN EXTENSIBLE

Dado un archivo con bloques con capacidad para 4 registros con el estado inicial que se especifica, desarrollar las transformaciones de estado sucesivas (describiendo actualizaciones en la tabla de dispersión y en bloques del archivo) al efectuar las siguientes operaciones:

- Alta de registro con clave 716(1011001100)
- Alta de registro con clave 232(11101000)
- Baja de registro con clave 798(1100011110)
- Baja de registro con clave 300(100101100)
- Baja de registro con clave 716(1011001100)
- Alta de registro con clave 103(1100111)
- Alta de registro con clave 175(10101111)

# DISPERSIÓN EXTENSIBLE

## Tabla de dispersión

### Bits de dispersión 3

0(000): 1

1(001): 2

2(010): 1

3(011): 3

4(100): 1

5(101): 2

6(110): 1

7(111): 0

## Archivo de datos

#Bloque	Bits	Clave R1	Clave R2	Clave R3	Clave R4
0	3	527 (1000001111)	743 (1011100111)	951 (1110110111)	
1	1	136 (10001000)	300 (100101100)	798 (1100011110)	832 (1101000000)
2	2	401 (110010001)	885 (1101110101)		
3	3	307 (100110011)	475 (111011011)	635 (1001111011)	

# ALTA DE REGISTRO CON CLAVE 716(1011001100)

## Tabla de dispersión

### Bits de dispersión 3

0(000): 4

1(001): 2

2(010): 1

3(011): 3

4(100): 4

5(101): 2

6(110): 1

7(111): 0

Tabla (100) = 1; overflow en bloque 1: bits de dispersión pasan de 1 a 2 y se agrega nueva cubeta. Como el bit de dispersión es menor al de la tabla no es necesario duplicar la tabla de direcciones. En la dirección del overflow apunto a la nueva cubeta y actualizo el puntero cada  $2^{\text{bits de dispersión}}$  de la cubeta, en este caso  $2^2$ , cada 4 posiciones. osea el 0 y el 4. Luego redisperso las claves entre las dos cubetas.

## Archivo de datos

#Bloque	Bits	Clave R1	Clave R2	Clave R3	Clave R4
0	3	527 (1000001111)	743 (1011100111)	951 (1110110111)	
1	2	798 (1100011110)			
2	2	401 (110010001)	885 (1101110101)		
3	3	307 (100110011)	475 (111011011)	635 (1001111011)	
4	2	136 (10001000)	300 (100101100)	716 (1011001100)	832 (1101000000)

# ALTA DE REGISTRO CON CLAVE 232(11101000)

## Tabla de dispersión

### Bits de dispersión 3

0(000): 5

1(001): 2

2(010): 1

3(011): 3

4(100): 4

5(101): 2

6(110): 1

7(111): 0

Tabla  $(000_2) = 4$ ; overflow en bloque 4: bits de dispersión pasan de 2 a 3, en registro 000 de la tabla se graba nuevo número de bloque 5 y se dispersa contenido en las cubetas

## Archivo de datos

#Bloque	Bits	Clave R1	Clave R2	Clave R3	Clave R4
0	3	527 (1000001111)	743 (1011100111)	951 (1110110111)	
1	2	798 (1100011110)			
2	2	401 (110010001)	885 (1101110101)		
3	3	307 (100110011)	475 (111011011)	635 (1001111011)	
4	3	300 (100101100)	716 (1011001100)		
5	3	136 (10001000)	232 (11101000)	832 (1101000000)	

# BAJA DE REGISTRO CON CLAVE 798(1100011110)

## Tabla de dispersión

### Bits de dispersión 3

0(000): 5

1(001): 2

2(010): 1

3(011): 3

4(100): 4

5(101): 2

6(110): 1

7(111): 0

798 = 1; bloque 1 queda vacío y tiene 2 bits de dispersión (se encuentra en los registros de la tabla que terminan con bits 10, y si en registros de la tabla terminados con bits 00 hay un mismo número de bloque se puede liberar el bloque 1 sustituyendo su número en la tabla por el número de bloque en registros terminados en 00: no se puede porque hay dos distintos). Se escribe el bloque 1 vacío.

## Archivo de datos

#Bloque	Bits	Clave R1	Clave R2	Clave R3	Clave R4
0	3	527 (1000001111)	743 (1011100111)	951 (1110110111)	
1	2	<del>798</del> <del>(1100011110)</del>			
2	2	401 (110010001)	885 (1101110101)		
3	3	307 (100110011)	475 (111011011)	635 (1001111011)	
4	3	300 (100101100)	716 (1011001100)		
5	3	136 (10001000)	232 (11101000)	832 (1101000000)	



# BAJA DE REGISTRO CON CLAVE 300(100101100)

Se ubica la clave y se borra

## Tabla de dispersión

### Bits de dispersión 3

0(000): 5

1(001): 2

2(010): 1

3(011): 3

4(100): 4

5(101): 2

6(110): 1

7(111): 0

## Archivo de datos

#Bloque	Bits	Clave R1	Clave R2	Clave R3	Clave R4
0	3	527 (1000001111)	743 (1011100111)	951 (1110110111)	
1	2				
2	2	401 (110010001)	885 (1101110101)		
3	3	307 (100110011)	475 (111011011)	635 (1001111011)	
4	3	716 (1011001100)			
5	3	136 (10001000)	232 (11101000)	832 (1101000000)	

# BAJA DE REGISTRO CON CLAVE 716(1011001100)

## Tabla de dispersión

### Bits de dispersión 3

0(000): 5

1(001): 2

2(010): 1

3(011): 3

4(100): 5

5(101): 2

6(110): 1

7(111): 0

Se vacía bloque 4 que tiene 3 bits de dispersión (se encuentra en el registro de la tabla que termina con bits 00; se puede reemplazar su número en la tabla por el número de bloque en registro 000, Tabla (0) = 5, restando 1 a los bits de dispersión del bloque 5). Se escribe bloque 5 por actualización de sus bits de dispersión y **se registra como libre el bloque 4.**

## Archivo de datos

#Bloque	Bits	Clave R1	Clave R2	Clave R3	Clave R4
0	3	527 (1000001111)	743 (1011100111)	951 (1110110111)	
1	2				
2	2	401 (110010001)	885 (1101110101)		
3	3	307 (100110011)	475 (111011011)	635 (1001111011)	
4	3	716 (1011001100)			
5	2	136 (10001000)	232 (11101000)	832 (1101000000)	

10

# ALTA DE REGISTRO CON CLAVE 103(1100111)

## Tabla de dispersión

### Bits de dispersión 3

0(000): 5

1(001): 2

2(010): 1

3(011): 3

4(100): 5

5(101): 2

6(110): 1

7(111): 0

## Archivo de datos

#Bloque	Bits	Clave R1	Clave R2	Clave R3	Clave R4
0	3	103 (1100111)	527 (1000001111)	743 (1011100111)	951 (1110110111)
1	2				
2	2	401 (110010001)	885 (1101110101)		
3	3	307 (100110011)	475 (111011011)	635 (1001111011)	
4	3	716 (1011001100)			
5	2	136 (10001000)	232 (11101000)	832 (1101000000)	

# ALTA DE REGISTRO CON CLAVE 175(10101111)

## Tabla de dispersión

### Bits de dispersión 4

0(0000): 5  
1(0001): 2  
2(0010): 1  
3(0011): 3  
4(0100): 5  
5(0101): 2  
6(0110): 1  
7(0111): 4  
8(1000): 5  
9(1001): 2  
2(1010): 1  
10(1011): 3  
11(1100): 5  
12(1101): 2  
13(1110): 1  
14(1111): 0

Tabla (111) = 0; overflow en bloque 0: bits de dispersión pasan de 3 a 4 y se agrega nueva cubeta (como hay uno libre se reutiliza el 4). Como el bit de dispersión es mayor al de la tabla es necesario duplicar la tabla de direcciones, se duplica la tabla repitiendo los punteros, en la dirección del overflow apunto a la nueva cubeta y luego redisperso las claves entre las dos cubetas.

## Archivo de datos

#Bloque	Bits	Clave R1	Clave R2	Clave R3	Clave R4
0	4	175 (10101111)	527 (1000001111)		
1	2				
2	2	401 (110010001)	885 (1101110101)		
3	3	307 (100110011)	475 (111011011)	635 (1001111011)	
4	4	103 (1100111)	743 (1011100111)	951 (1110110111)	
5	2	136 (10001000)	232 (11101000)	832 (1101000000)	