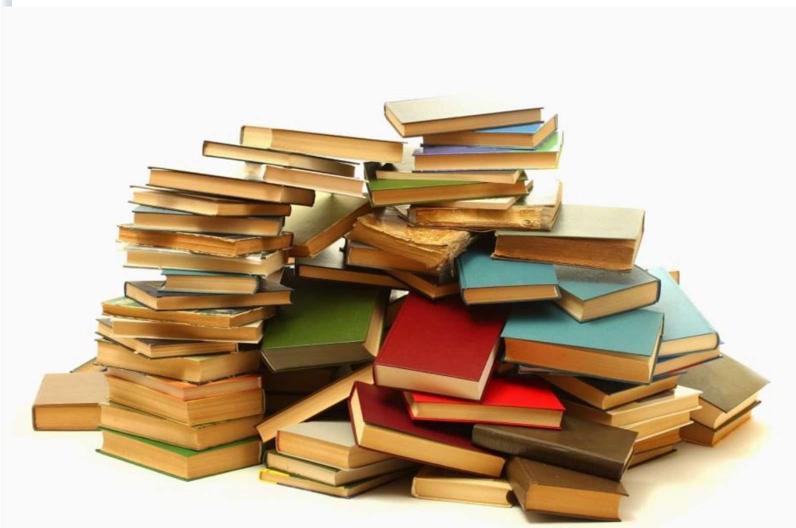
h) Familia de árboles B

- Traer y llevar bloques de información desde un disco de memoria secundaria a la memoria principal para que la CPU opere, es muy costoso en términos de tiempo.

$$\frac{Disco}{Memoria\ Principal} = \frac{50\ mseg}{500\ nseg} = \frac{50 * 10^{-3} seg}{500 * 10^{-9} seg} = \frac{\mathbf{100.000}}{\mathbf{1}}$$

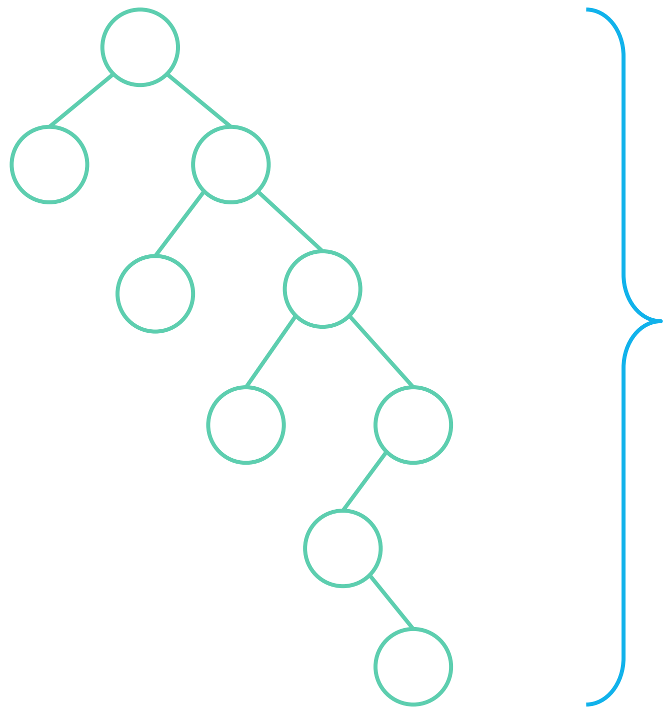
h) Familia de árboles B

- Por lo tanto, para mejorar la eficiencia de las operaciones, los datos deben ser organizados de forma que se **minimice la cantidad de accesos al disco** (operaciones E/S **O.A.**).



h) Familia de árboles B

- Entonces ya no son tan buenos los AB ($O(\log_2 n)$) porque son árboles **flacos y altos**.



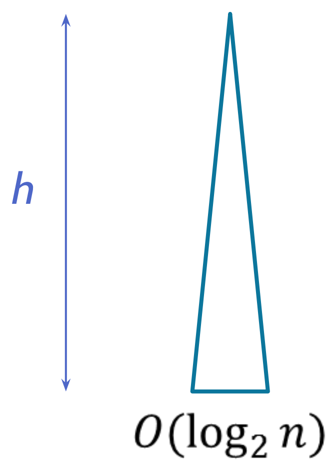
Muchos niveles
Muchas comparaciones
Datos dispersos

h) Familia de árboles B

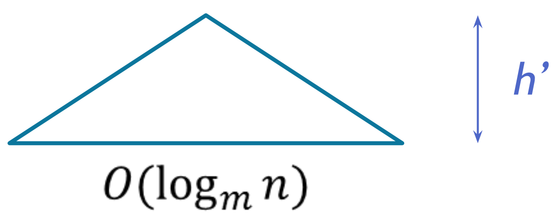
- <u>Solución</u>: **Árbol M-ario**

  - Se «*aplana*» el árbol ya que sus nodos pueden almacenar más de un dato y tener más de 2 hijos.

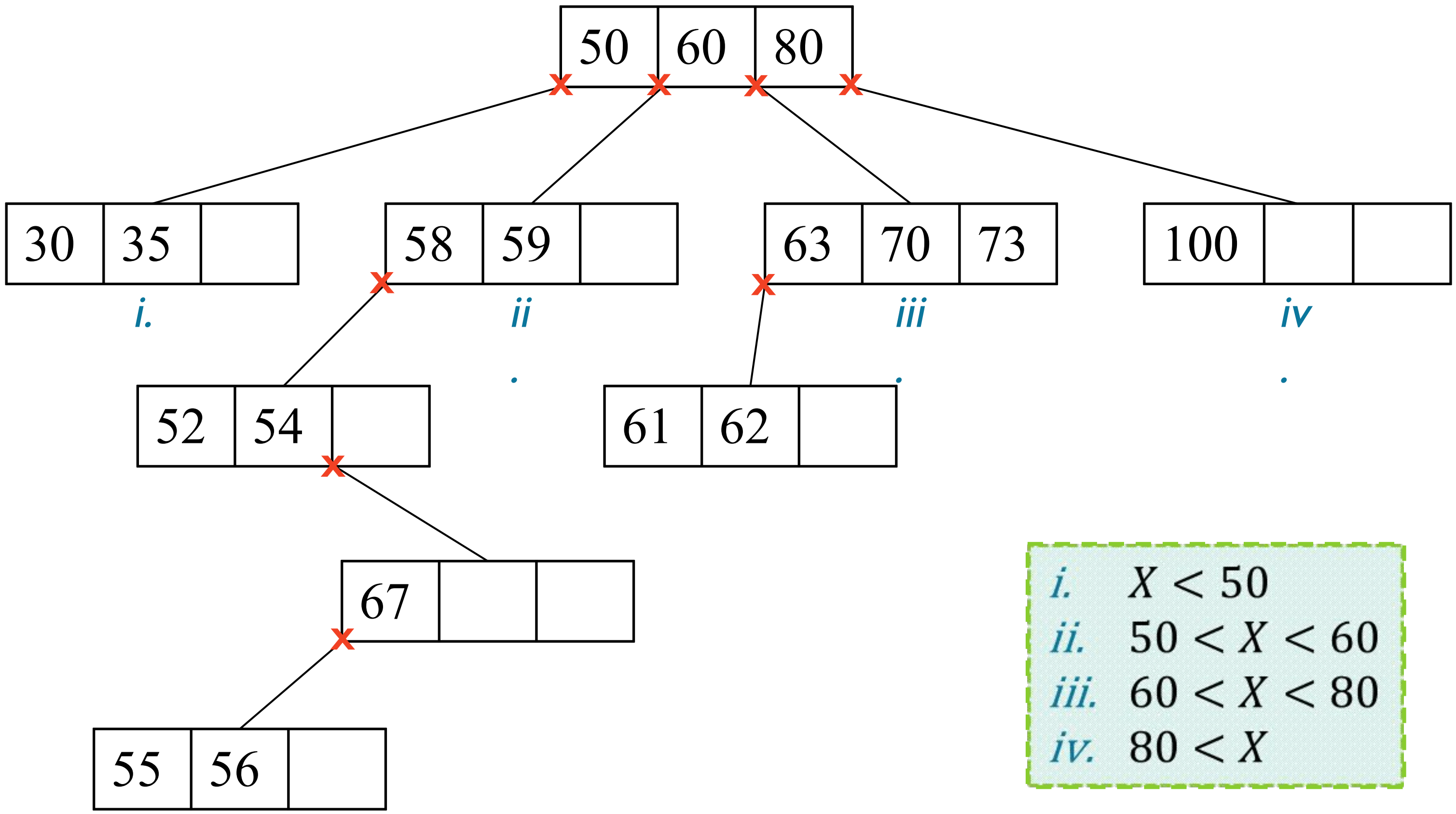
<u>BINARIO (ABB)</u>



<u>M-ARIO (AM-arioB)</u>



h) Familia de árboles B



i. $X < 50$

ii. $50 < X < 60$

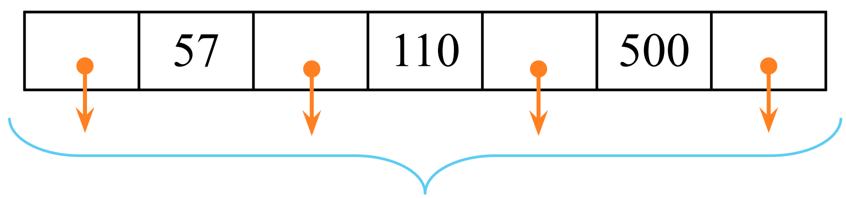
iii. $60 < X < 80$

iv. $80 < X$

h) Familia de árboles B

- En el ejemplo anterior: $\boxed{M = 4}$
  - Cantidad máxima de hijos = M = 4 hijos
  - Cantidad máxima de claves = M-1 = 3 claves

- Los nodos tienen la siguiente forma:

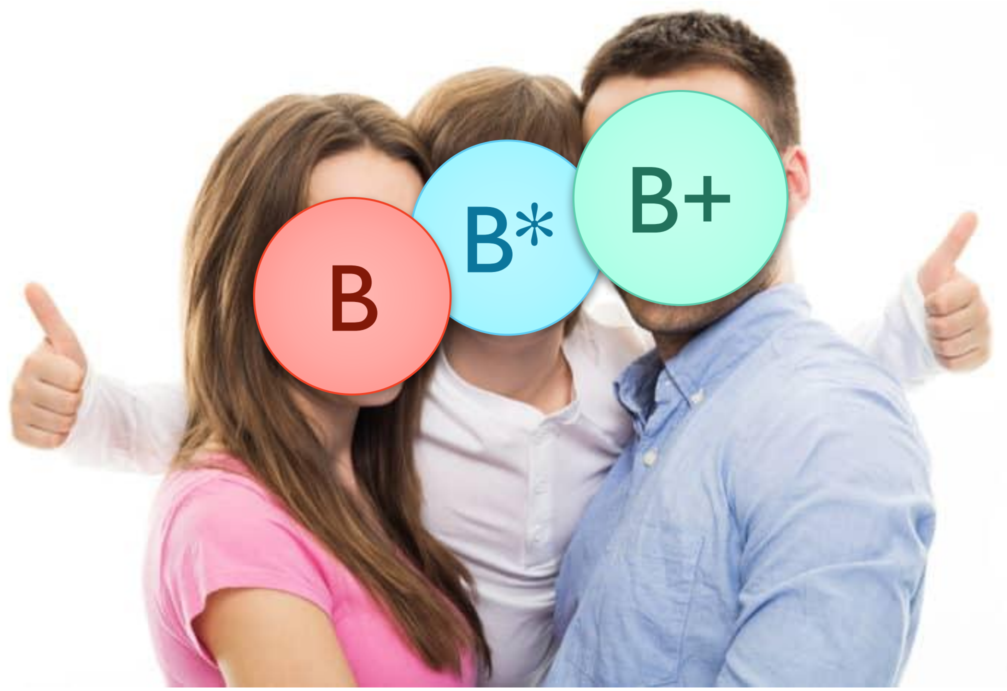


4 punteros, 3 claves

- Este árbol también se puede desbalancear y aumentar su *h*, de aquí nacen los **árboles B**.

h) Familia de árboles B

- Dentro de la «*Familia de árboles B*» veremos:



h) Familia de árboles B (B)

Árbol B

- Totalmente balanceado.
- Un árbol B de orden M es un árbol de búsqueda multi-senda que puede estar vacío o satisfacer las siguientes propiedades:

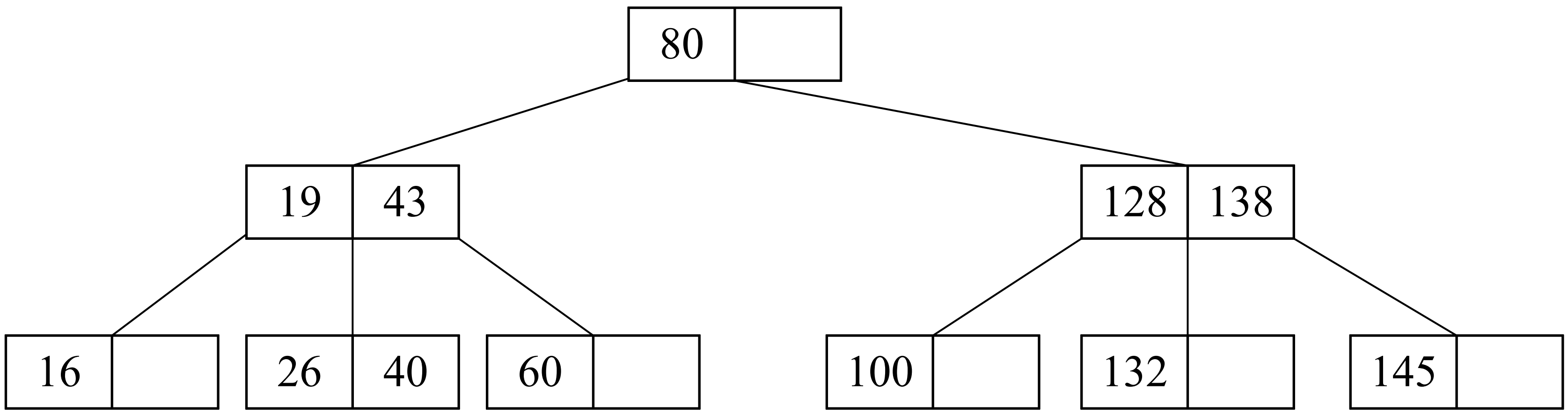
1
- La raíz tiene al menos 2 hijos.

2
- Todos los nodos distintos al nodo raíz y a las hojas, tienen como mínimo hijos.

3
- Todas las hojas están al mismo nivel.

h) Familia de árboles B (B)

- Ejemplo de árbol B con M = 3



En el peor caso, un árbol B está ½ lleno.

h) Familia de árboles B (B)

- **Inserción**
  - Ejemplo: M = 4

- $\boxed{Cantidad\ máx.\ hijos = M} = 4$ ➡ $\boxed{Cantidad\ máx.\ claves = M - 1} = 3$
- $\boxed{Cantidad\ mín.\ hijos = \left\lceil \frac{M}{2} \right\rceil} = 2$ ➡ $\boxed{Cantidad\ mín.\ claves = \left\lceil \frac{M}{2} \right\rceil - 1} = 1$

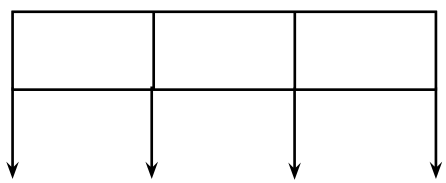
<u>SIEMPRE</u> se deben calcular estos 4 valores antes de empezar a trabajar con un árbol B

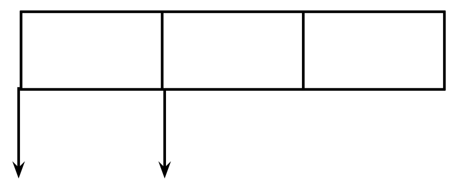
h) Familia de árboles B (B)

M = 4

- $Cantidad\ máx.\ hijos = 4$ ➡ $Cantidad\ máx.\ claves = 3$
- $Cantidad\ mín.\ hijos = 2$ ➡ $Cantidad\ mín.\ claves = 1$

Sabiendo esto, podemos deducir que sus nodos tienen la siguiente forma:





h) Familia de árboles B (B)

+50	50		
+60	50	60	
+1	1	50	60

h) Familia de árboles B (B)

+30

$$\overbrace{\begin{array}{|c|c|c|}\hline 1 & 50 & 60 \\ \hline\end{array}}^{¿¿30??}$$

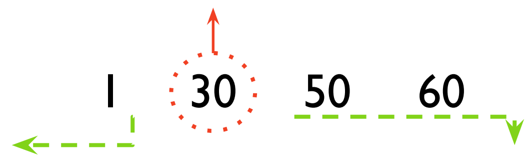
No hay espacio para el 30, esto provoca un ***Overflow***

1° Se elige el «dato del medio»

$$\underbrace{1 \quad 30 \quad 50 \quad 60}$$

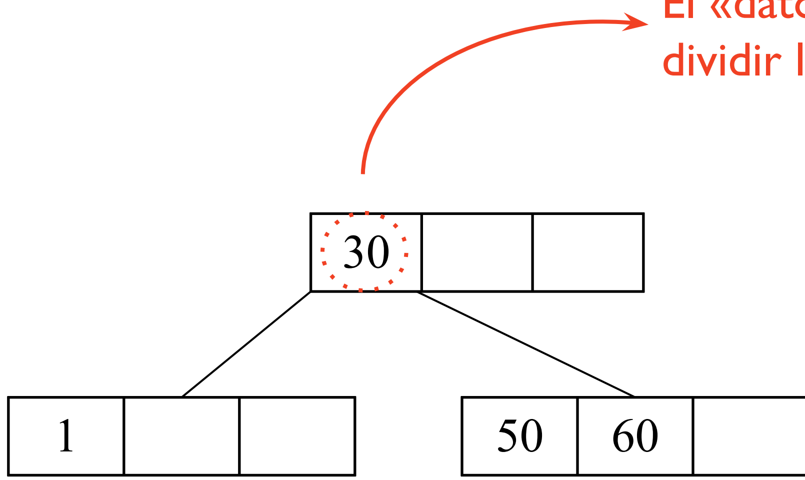
$$n = 4 \Rightarrow \left\lceil \frac{n}{2} \right\rceil = 2$$

h) Familia de árboles B (B)

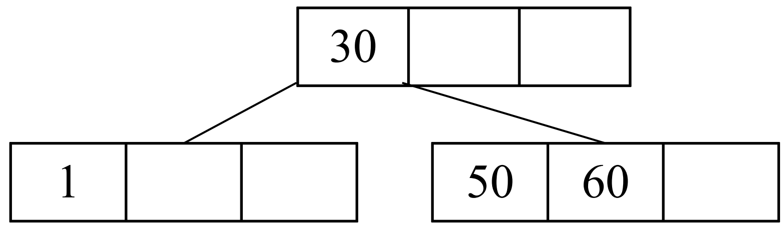


2° Se divide el nodo

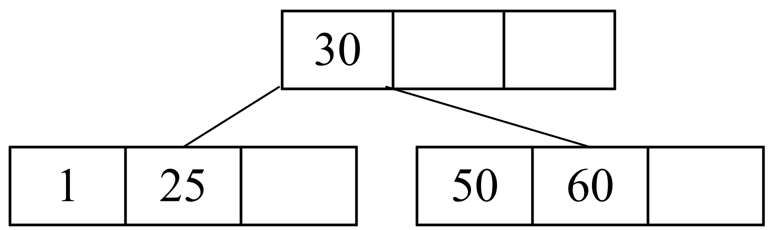
El «dato del medio» subió para dividir los nodos



h) Familia de árboles B (B)

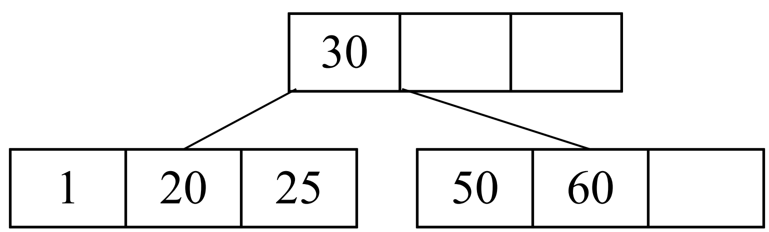


+25

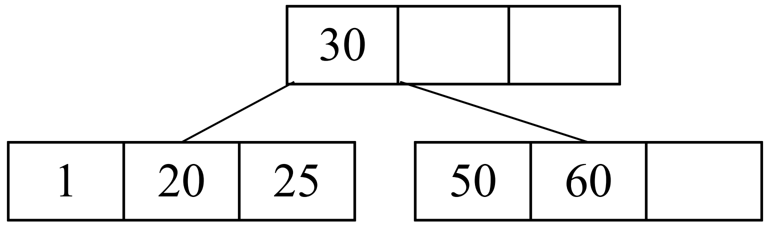


[Siempre se busca insertar en las hojas]

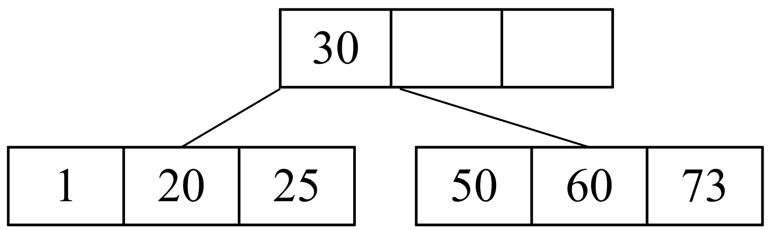
+20



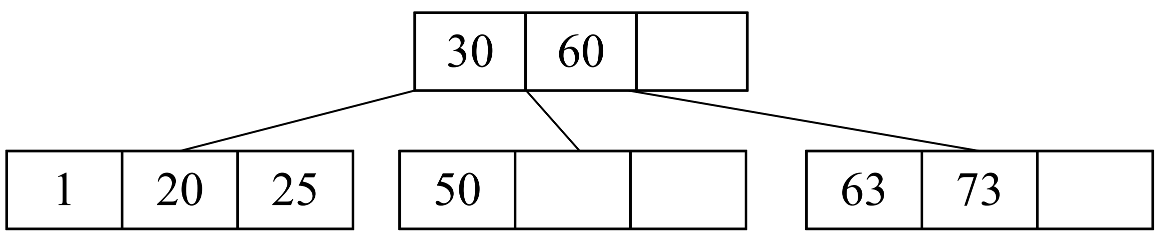
h) Familia de árboles B (B)



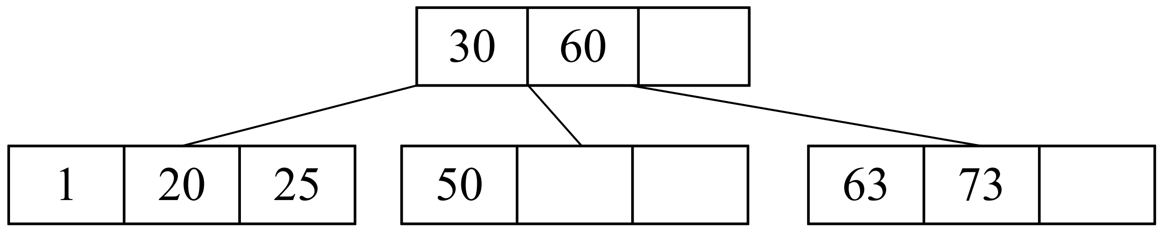
+73



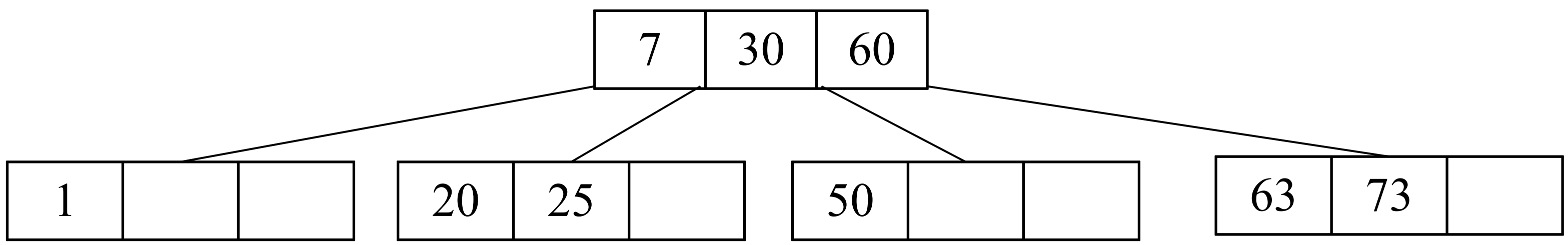
+63 **Overflow** □ 50 60 63 73



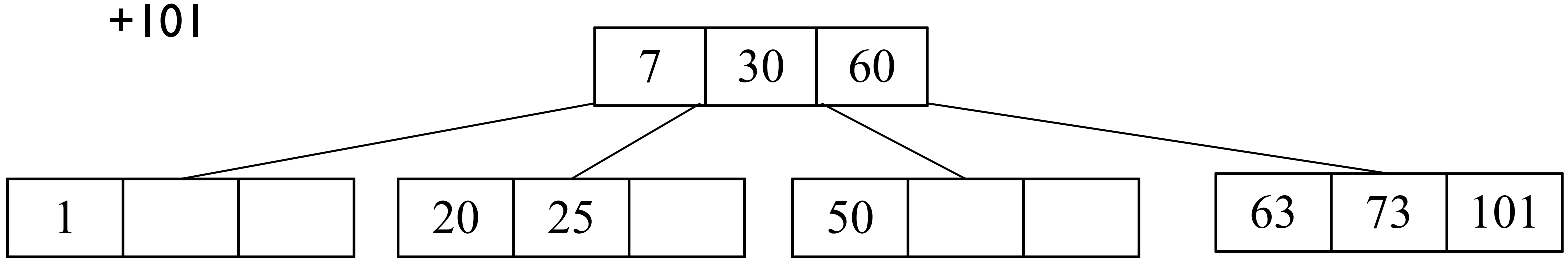
h) Familia de árboles B (B)



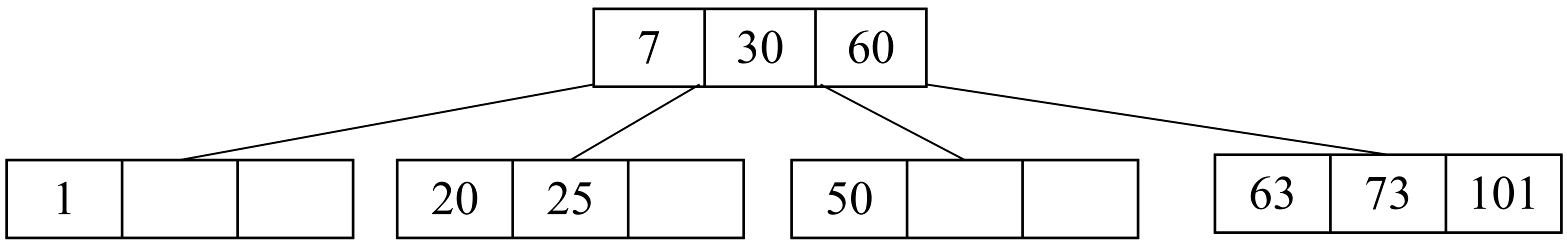
+7 ***Overflow*** □ 1 7 20 25



+101

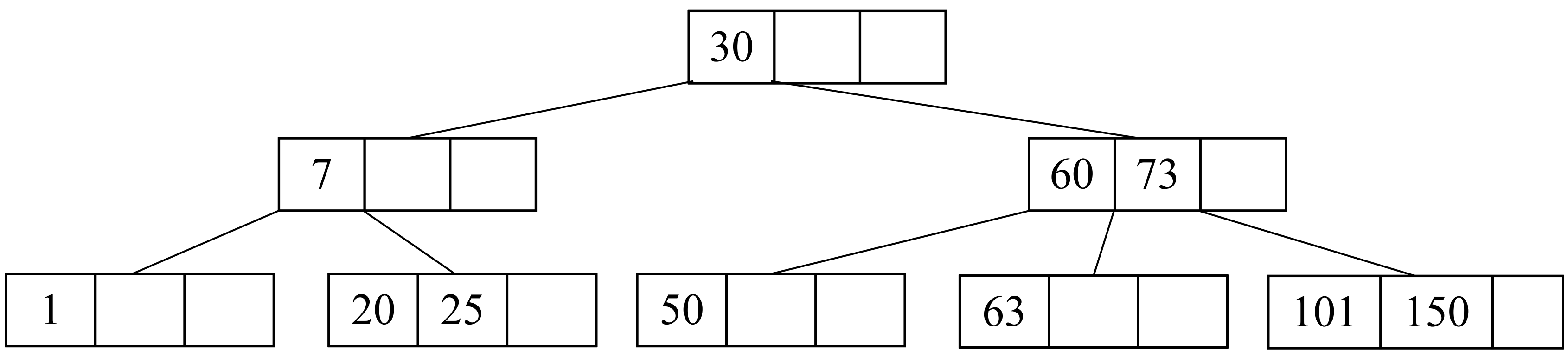


h) Familia de árboles B (B)



+150 ***Overflow*** □ 63 73 101 150

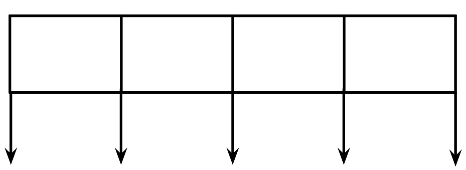
Overflow □ 7 30 60 73

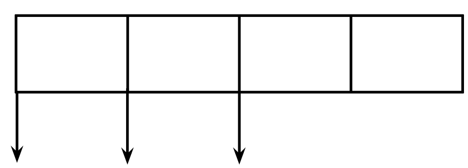


h) Familia de árboles B (B)

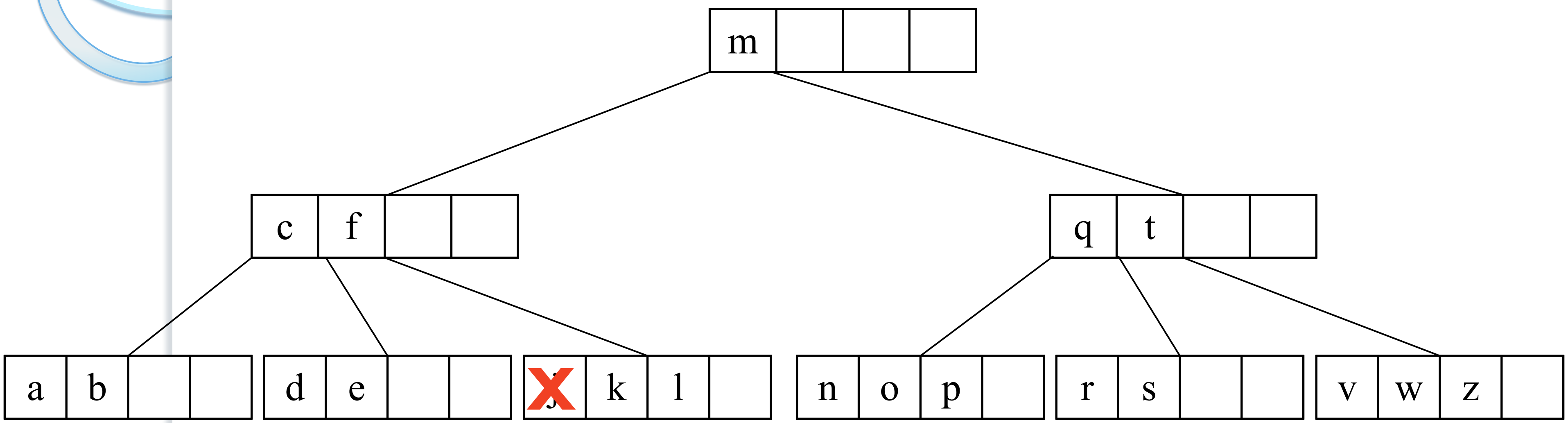
- **Eliminación**
  - Ejemplo: M = 5

- $\boxed{\text{Cantidad máx. hijos} = M} = 5$ → $\boxed{\text{Cantidad máx. claves} = M - 1} = 4$
- $\boxed{\text{Cantidad mín. hijos} = \left\lceil \frac{M}{2} \right\rceil} = 3$ → $\boxed{\text{Cantidad mín. claves} = \left\lceil \frac{M}{2} \right\rceil - 1} = 2$



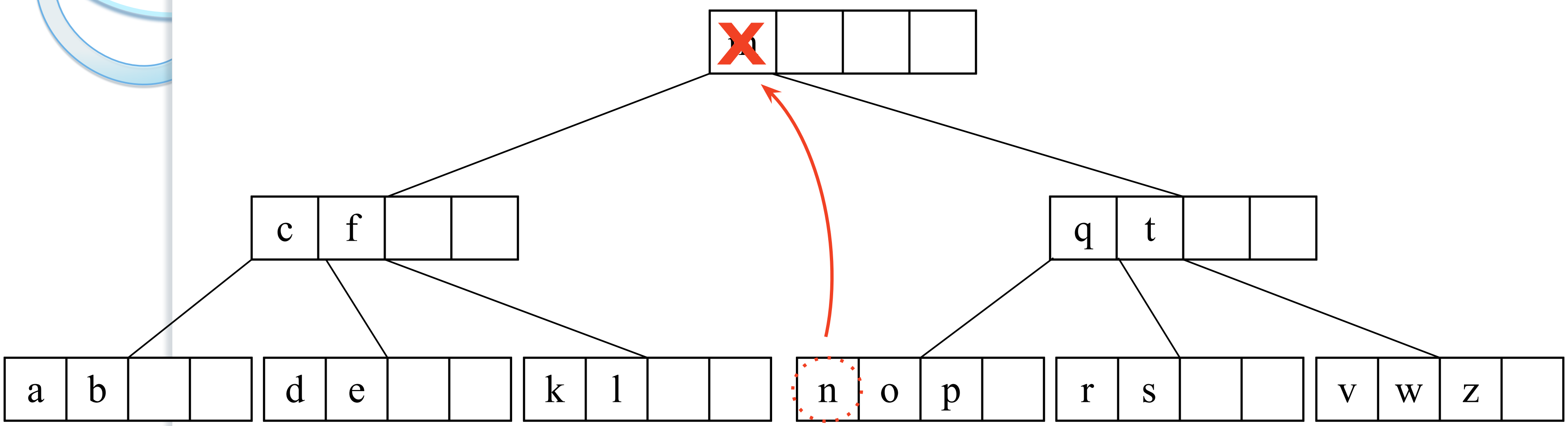


h) Familia de árboles B (B)



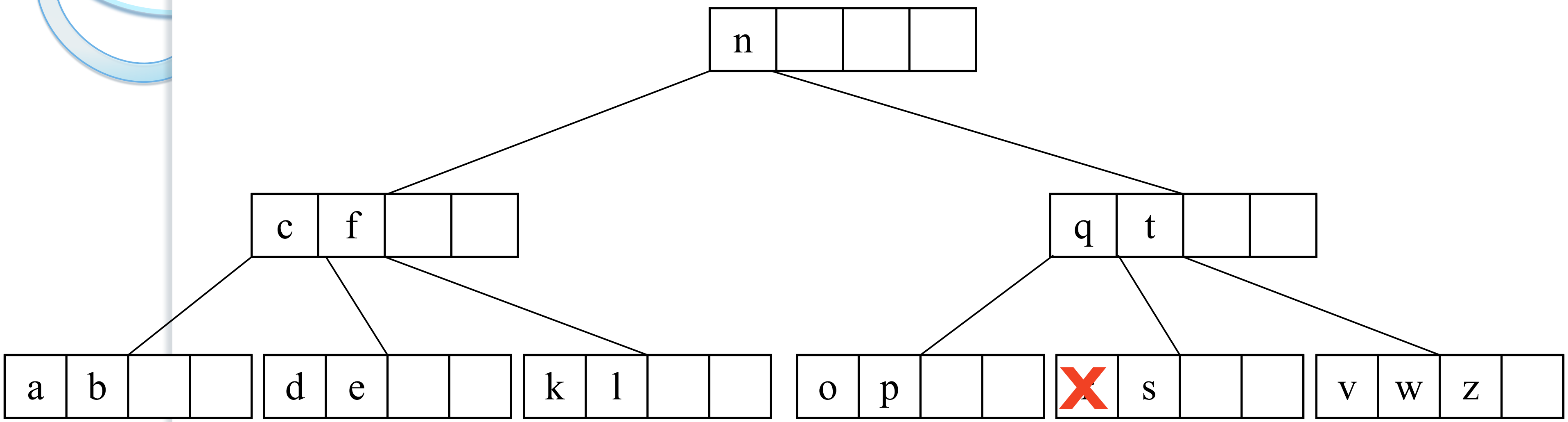
-j Se puede eliminar sin problema

h) Familia de árboles B (B)



-m Cuando se elimina un valor que no está en las hojas, sube el sucesor inmediato. En este caso el sucesor inmediato de *m*, es *n*.

h) Familia de árboles B (B)

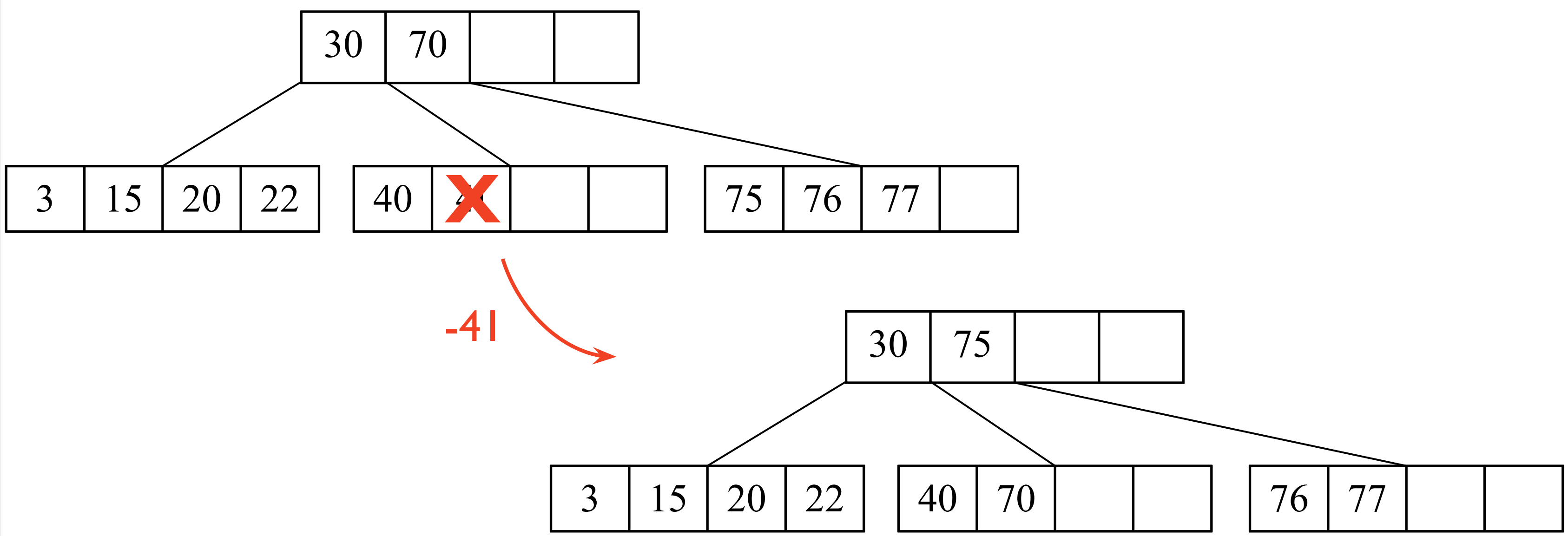


-r Va a quedar con menos de dos claves, esto provoca un ***Underflow***

h) Familia de árboles B (B)

Cuando hay ***Underflow*** hacer:

1. **Redistribución** *(si el hermano puede, le pasa uno)*
   Primero se ve el hermano derecho, luego el izquierdo.
   El separador de los hermanos baja y el sobrante del hermano sube.



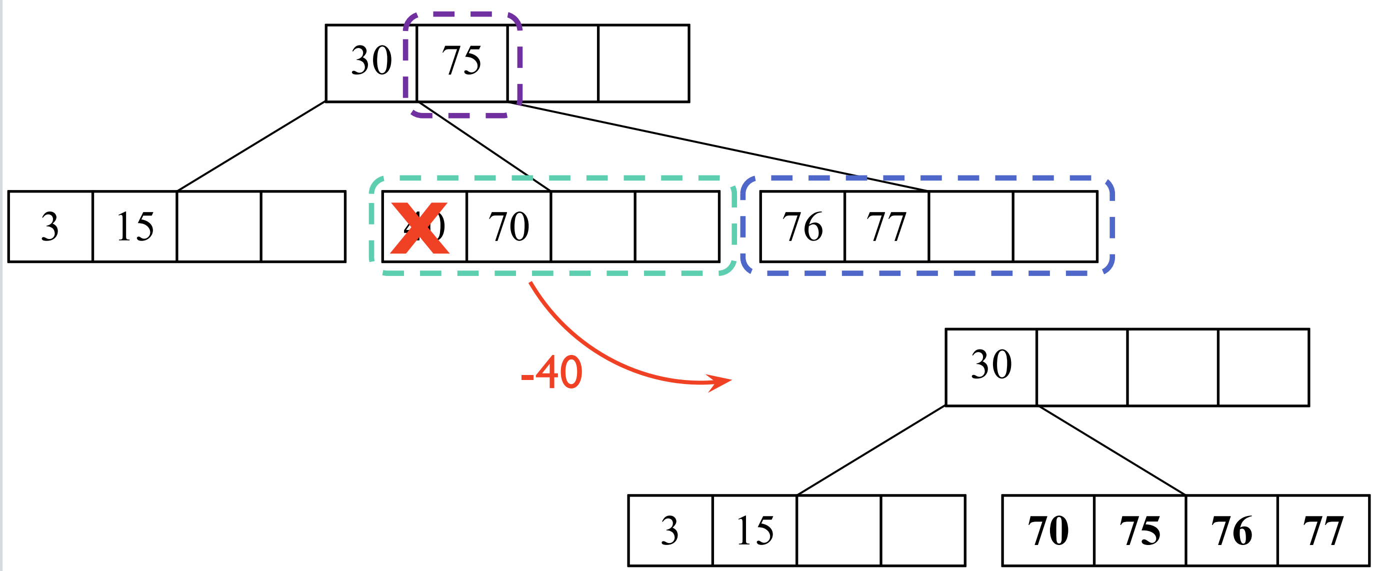
h) Familia de árboles B (B)

2. **Mezcla** *(cuando ninguno de los hermanos le puede dar)*

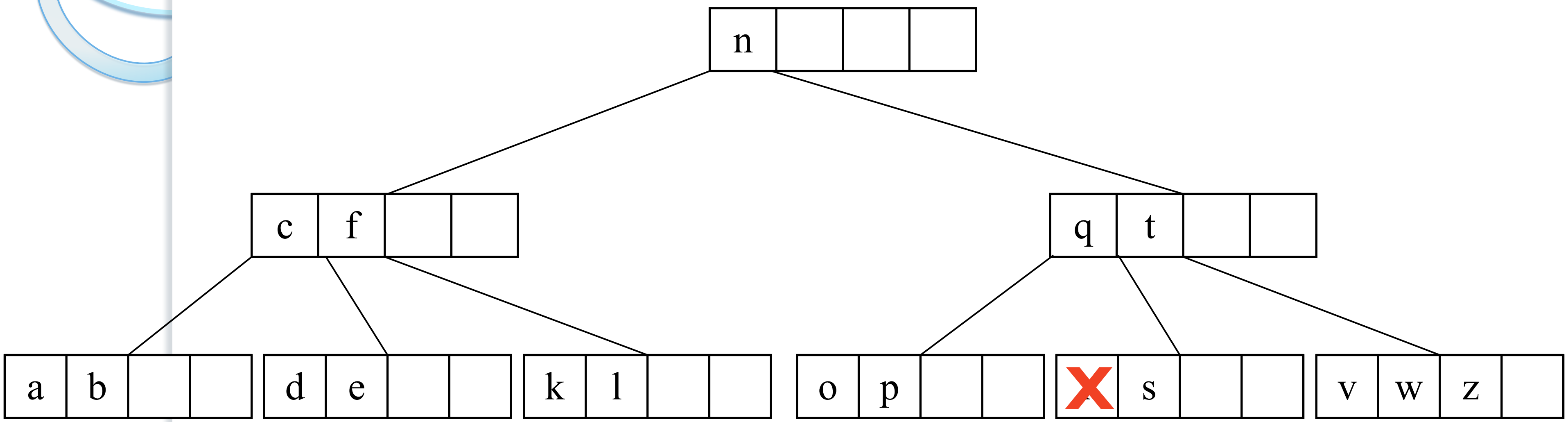
Se junta: + Nodo problema (el de la insuficiencia)

+ Separador

+ Hermano derecho (si no tiene, entonces izquierdo)

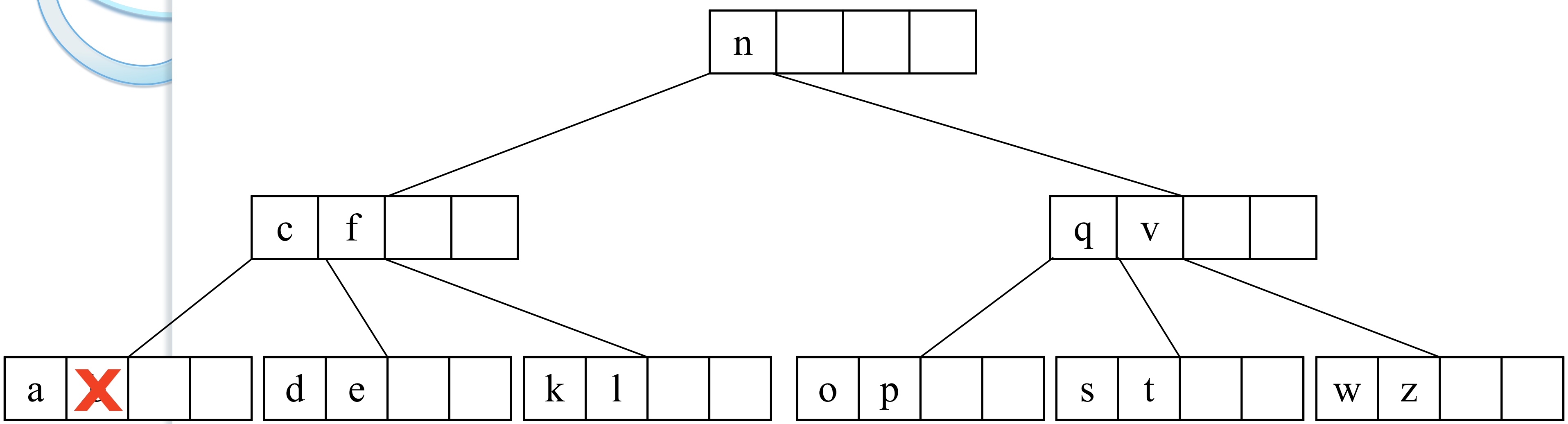


h) Familia de árboles B (B)



-r *Underflow* □ 1° ¿Redistribuir?

h) Familia de árboles B (B)



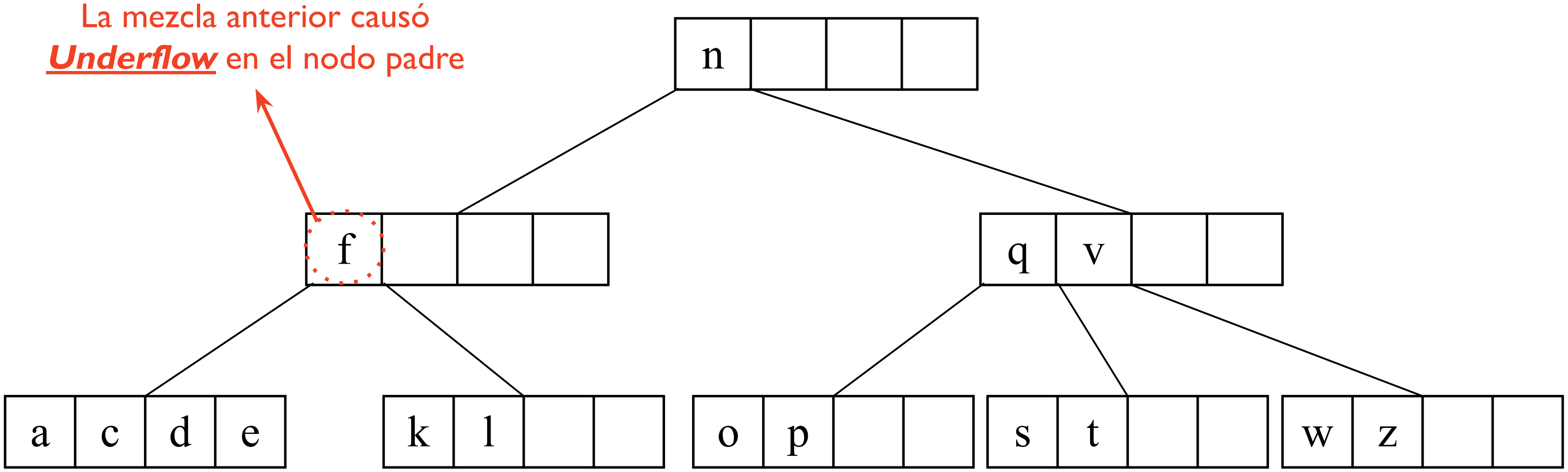
-b *Underflow* □ 1° ¿Redistribuir?

2° Mezclar: a c d e

N. Prob *Sep* *Hermano Der*

h) Familia de árboles B (B)

La mezcla anterior causó
<u>Underflow</u> en el nodo padre

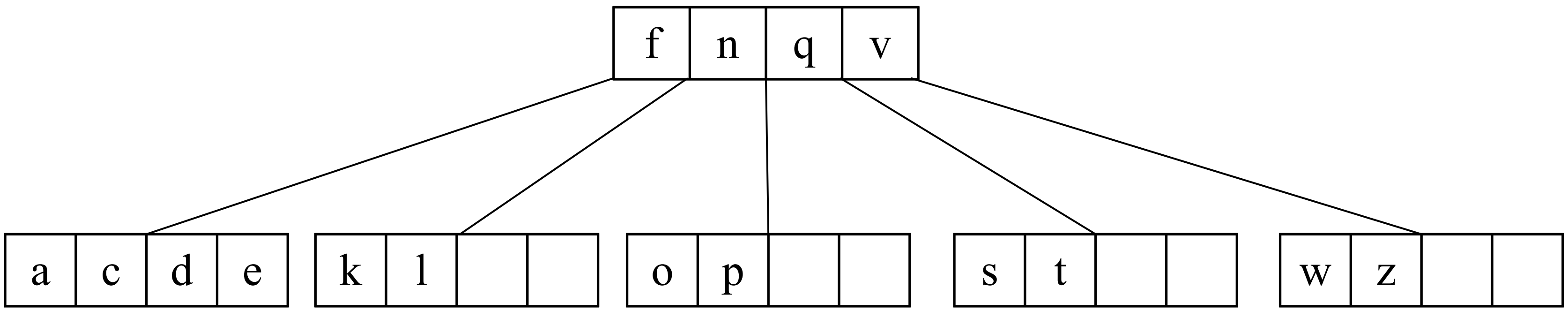


<u>Underflow</u> □ 1° ¿Redistribuir? ✗

2° Mezclar: f n q v

N. Prob *Sep* *Hermano Der*

h) Familia de árboles B (B)

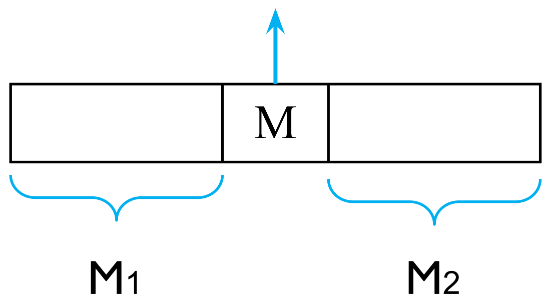


h) Familia de árboles B (B*)

Árbol B*

- Es una variación de un árbol B.
- La idea básica es reducir la frecuencia de la división de nodos.
- Cuando un nodo se llena, en lugar de efectuar una división, se realiza un proceso de **redistribución**.
- Si se necesita efectuar división de nodos, dos nodos son divididos para formar 3 nodos, los que están aprox. 2/3 llenos.

h) Familia de árboles B (B*)



- B
- B*

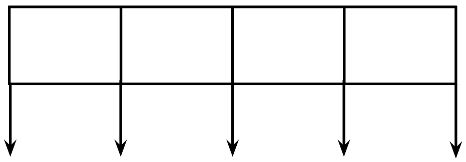
1° Redistribución
2° División

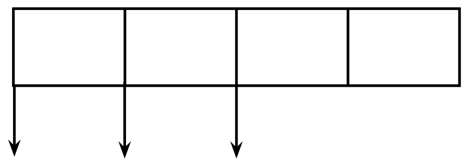
h) Familia de árboles B (B*)

- Ej. Redistribución

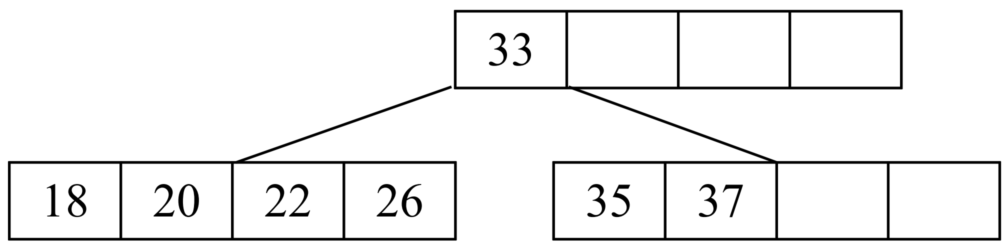
M = 5

- $\boxed{Cantidad\ máx.\ hijos = M} = 5$ → $\boxed{Cantidad\ máx.\ claves = M - 1} = 4$
- $\boxed{Cantidad\ mín.\ hijos = \left\lceil \frac{M}{2} \right\rceil} = 3$ → $\boxed{Cantidad\ mín.\ claves = \left\lceil \frac{M}{2} \right\rceil - 1} = 2$





h) Familia de árboles B (B*)

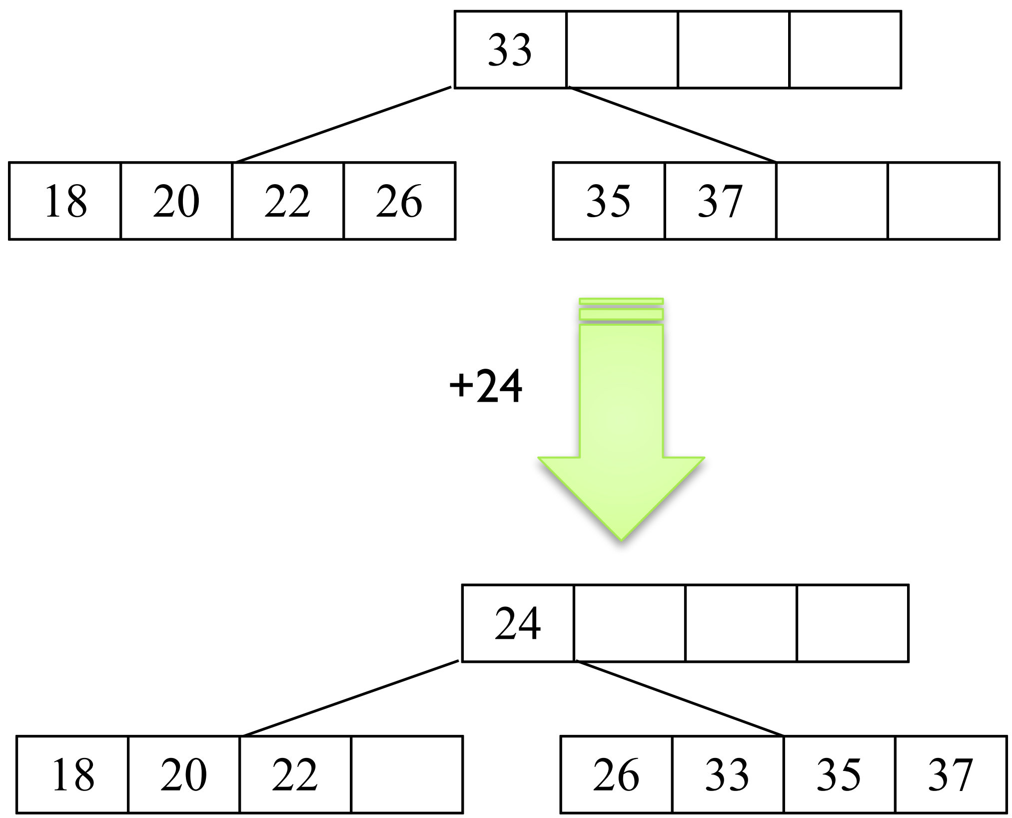


+24 ***Overflow*** □ 1° ¿Redistribución?

$$\underbrace{\underbrace{18 \quad 20 \quad 22 \quad \mathbf{24} \quad 26}_{\textit{N. Prob}} \quad \underbrace{33}_{\textit{Sep}} \quad \underbrace{35 \quad 37}_{\textit{Hermano Der}}}$$

$$\left\lceil \frac{n}{2} \right\rceil = \left\lceil \frac{8}{2} \right\rceil = 4\ (sube)$$

h) Familia de árboles B (B*)

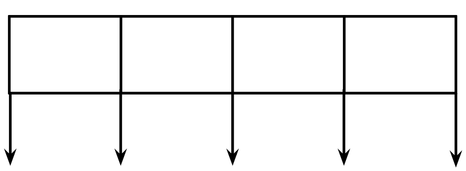


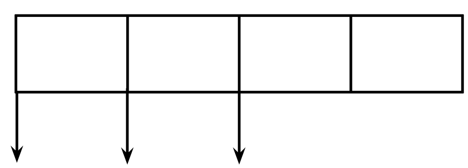
h) Familia de árboles B (B*)

- Ej. División

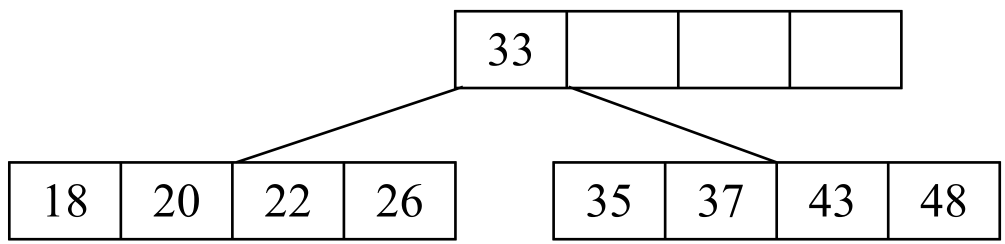
M = 5

- $\boxed{Cantidad\ máx.\ hijos = M} = 5$ ➡ $\boxed{Cantidad\ máx.\ claves = M - 1} = 4$
- $\boxed{Cantidad\ mín.\ hijos = \left\lceil \frac{M}{2} \right\rceil} = 3$ ➡ $\boxed{Cantidad\ mín.\ claves = \left\lceil \frac{M}{2} \right\rceil - 1} = 2$





h) Familia de árboles B (B*)



+24 **Overflow** □ 1° ¿Redistribución?
2° División

$$\underbrace{18 \quad 20 \quad 22 \quad \mathbf{24} \quad 26}_{\textit{N. Prob}} \quad \underbrace{33}_{\textit{Sep}} \quad \underbrace{35 \quad 37 \quad 43 \quad 48}_{\textit{Hermano Der}}$$

Hay que obtener 3 nodos, ¿dónde hacemos las divisiones?

h) Familia de árboles B (B*)

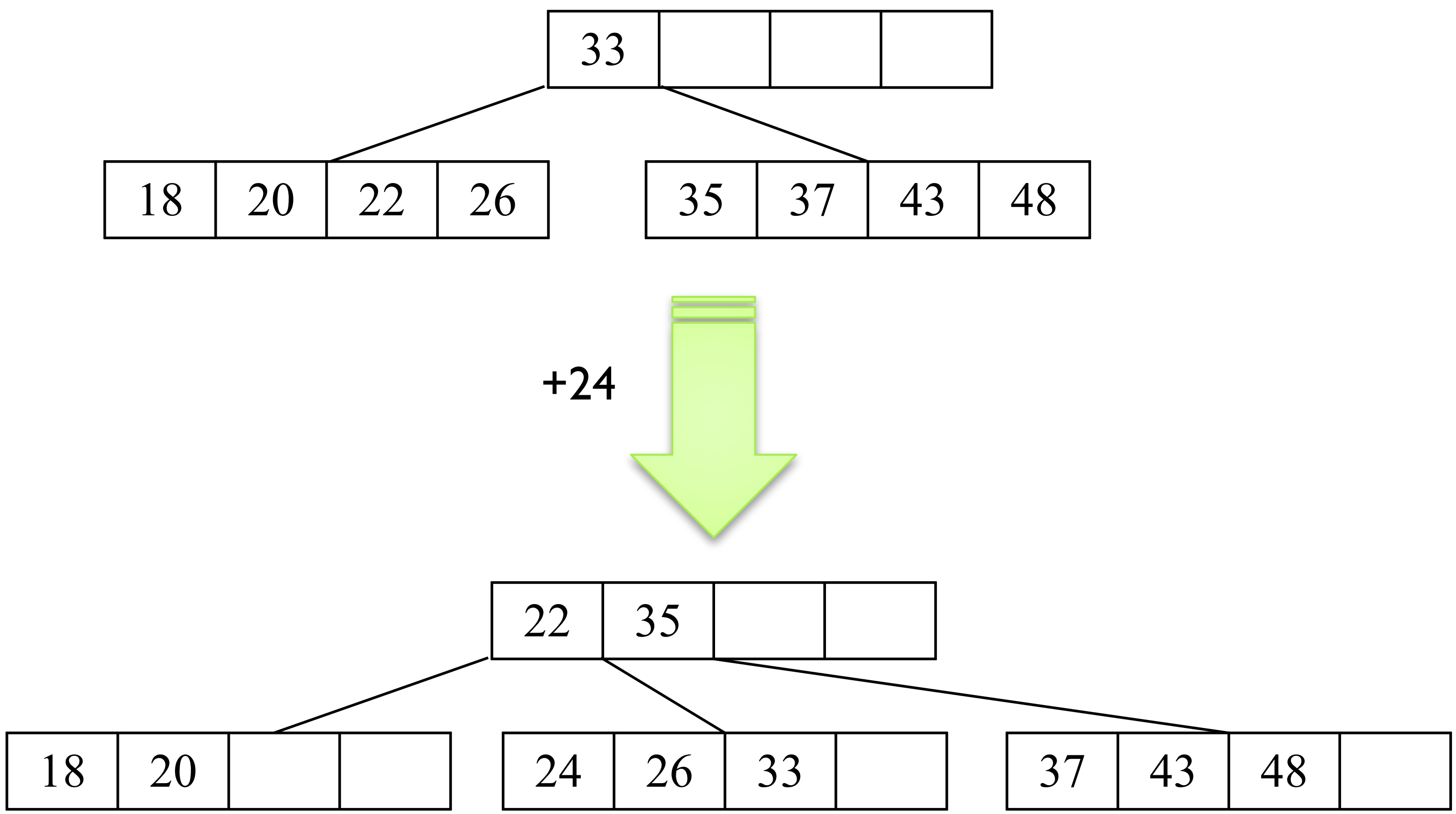
- Las divisiones se hacen después de:

$$\boxed{\alpha = \left\lfloor \frac{2M-2}{3} \right\rfloor} = \left\lfloor \frac{2*5-2}{3} \right\rfloor = \left\lfloor \frac{8}{3} \right\rfloor = 2$$

$$\boxed{\beta = \left\lfloor 2 * \frac{2M-2}{3} \right\rfloor + 1} = \left\lfloor 2 * \frac{2*5-2}{3} \right\rfloor + 1 = \left\lfloor 2 * \frac{8}{3} \right\rfloor + 1 = 5 + 1 = 6$$

	α				β				
18	20	22	**24**	26	33	35	37	43	48

h) Familia de árboles B (B*)



h) Familia de árboles B (B+)

Árbol B+

- Es otra variación de un árbol B.
- Su característica principal es que contiene todos los elementos del árbol en el nivel de las hojas.
- Los nodos internos contienen claves para guiar la búsqueda.
- Las hojas se unen en una lista lineal, permitiendo accesos secuenciales eficientes a los datos. Por lo tanto, la estructura de los nodos internos difiere de la de los nodos hoja.

h) Familia de árboles B (B+)

- Diremos entonces que un árbol B+ tiene 2 partes:

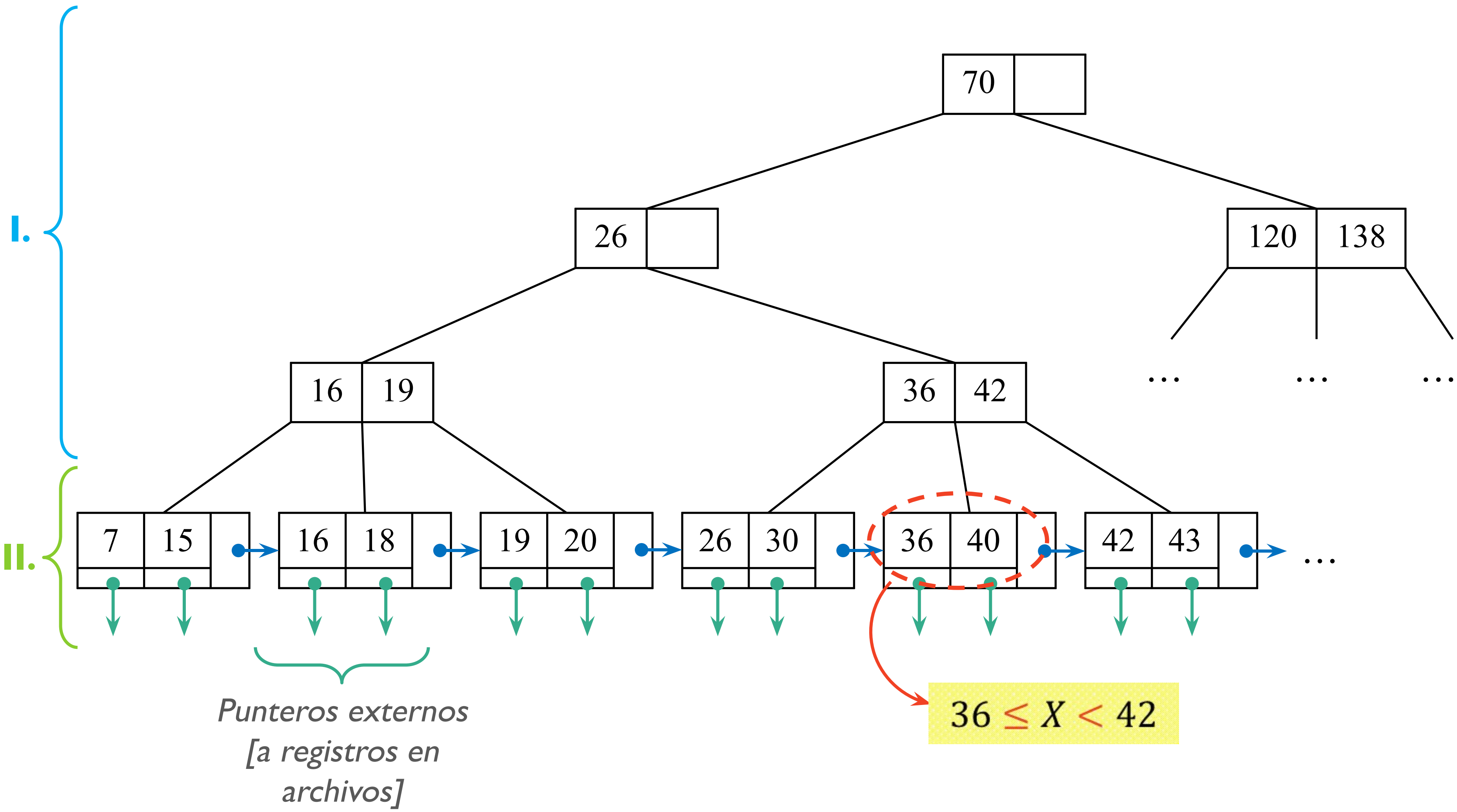
I. Parte Índice

- Nodos internos.
- Se comporta como un árbol B.

II. Conjunto Secuencia

- Nodos hoja.
- Están todos al mismo nivel, unidos por punteros en una lista lineal.

- Ejemplo de árbol B+:



h) Familia de árboles B (B+)

- Para buscar un dato específico se usa el árbol B de **I.**
- Para un recorrido secuencial se utiliza **II.**

- Estos árboles tienen M y M_{HOJA}

I) Índice

$$Cant.\ máx\ \textbf{hijos} = M$$

$$Cant.\ máx\ claves = M - 1$$

$$Cant.\ min\ hijos = \left\lceil \frac{M}{2} \right\rceil$$

$$Cant.\ min\ claves = \left\lceil \frac{M}{2} \right\rceil - 1$$

II) Conj. Secuencia

$$Cant.\ máx\ \textbf{claves} = M_{HOJA}$$

$$Cant.\ min\ claves = \left\lceil \frac{M_{HOJA}}{2} \right\rceil$$

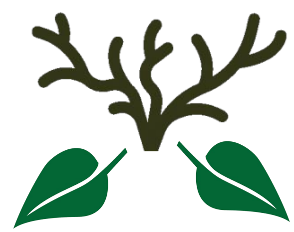
Si solo se da un valor de M, se asume que el valor de M_{HOJA} es el mismo.

h) Familia de árboles B (B+)

- **Inserción**
  - Cuando la raíz es el único nodo en el árbol, se consideran nodos tipo hoja.

M_{HOJA}

  - Cuando se crea más de un nivel, el árbol se divide en nodos internos y hojas.



M

M_{HOJA}

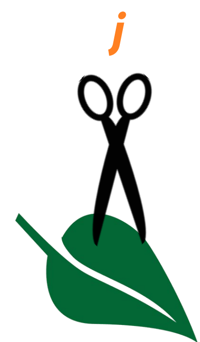
  - ***TODAS*** las claves están en las hojas.

h) Familia de árboles B (B+)

- Cuando se genera un *overflow* hay que dividir el nodo.

- El corte se hace en el dato:

$$j = \left\lceil \frac{M_{HOJA} + 1}{2} \right\rceil$$



- El valor en «j+1» se repite en el nodo padre.

- Si hay *overflow* en los nodos internos, se trata como en un árbol B normal.

h) Familia de árboles B (B+)

- Ejemplo: Insertar en un árbol B+ con M=3

  I) Índice:

$$Cant.\ máx\ hijos = M = 3 \Rightarrow Cant.\ máx\ claves = M - 1 = 2$$

$$Cant.\ min\ hijos = \left\lceil \frac{M}{2} \right\rceil = 2 \Rightarrow Cant.\ min\ claves = \left\lceil \frac{M}{2} \right\rceil - 1 = 1$$

  II) Conjunto secuencia:

$$Cant.\ máx\ claves = M_{HOJA} = 3$$

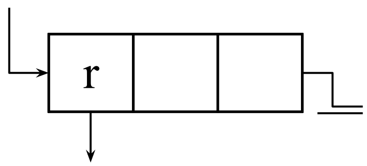
$$Cant.\ min\ claves = \left\lceil \frac{M_{HOJA}}{2} \right\rceil = 2$$

$$j = \left\lceil \frac{M_{HOJA} + 1}{2} \right\rceil = \left\lceil \frac{3+1}{2} \right\rceil = 2$$

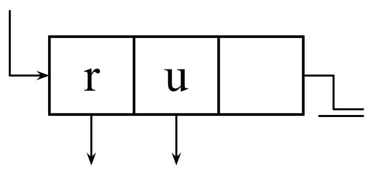
SIEMPRE se deben calcular estos 7 valores antes de empezar a trabajar con un árbol B+

h) Familia de árboles B (B+)

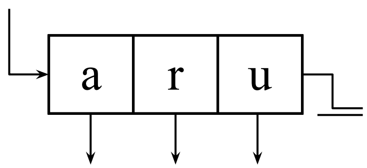
+r



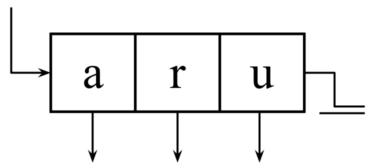
+u



+a



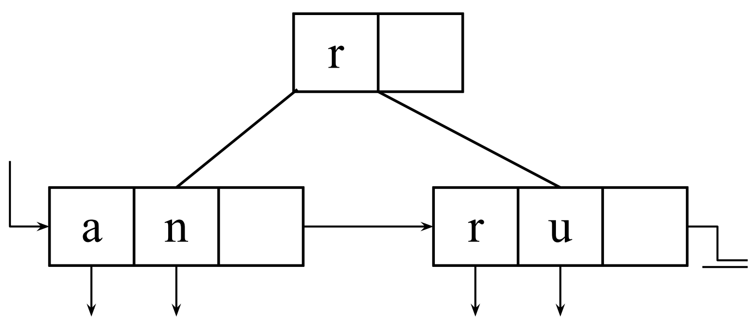
h) Familia de árboles B (B+)



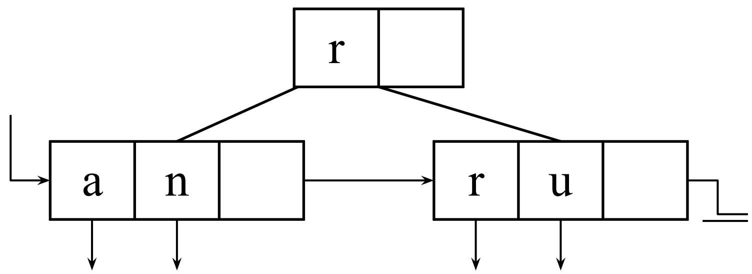
+n **<u>*Overflow*</u>** □ Dividir: «j» entradas en el nodo original.

«j+1» se copia arriba

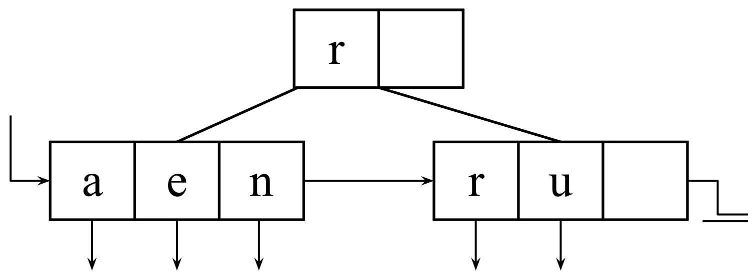
a n r u



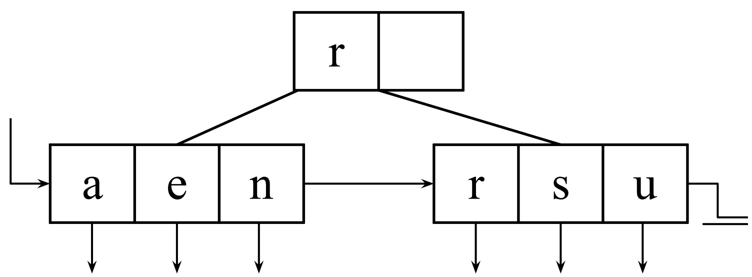
h) Familia de árboles B (B+)



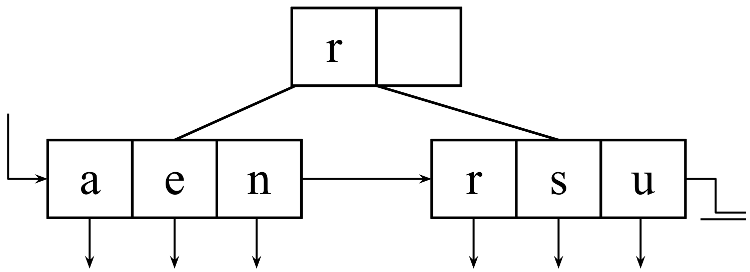
+e



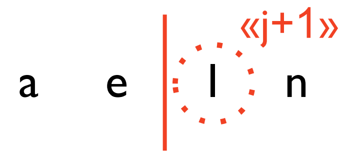
+s

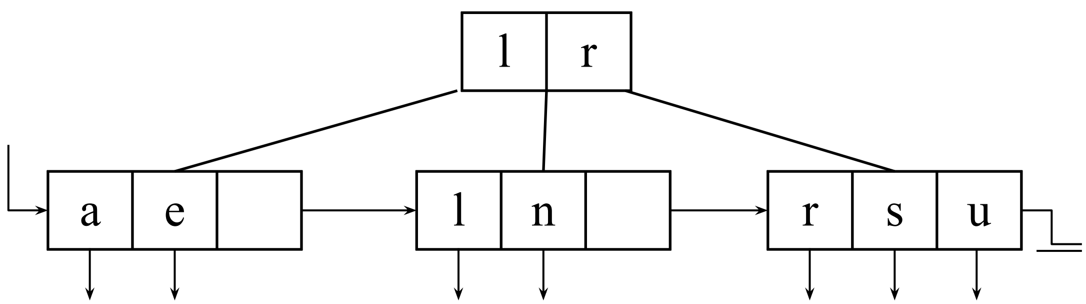


h) Familia de árboles B (B+)

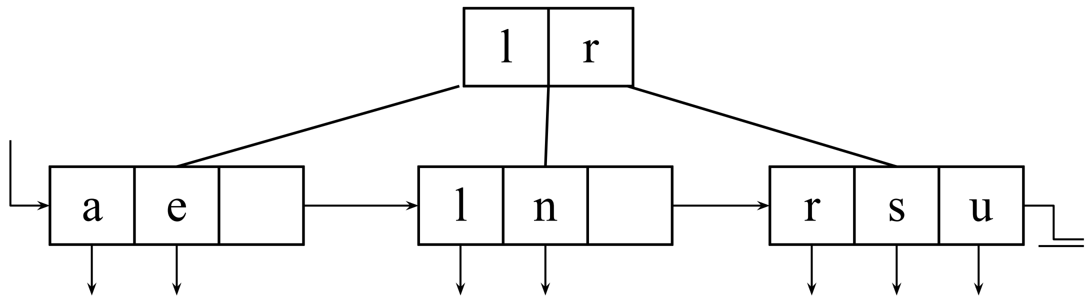


+l **Overflow** □ Dividir: «j» entradas en el nodo original.





h) Familia de árboles B (B+)



+v ***Overflow*** □ Dividir: «j» entradas en el nodo original.

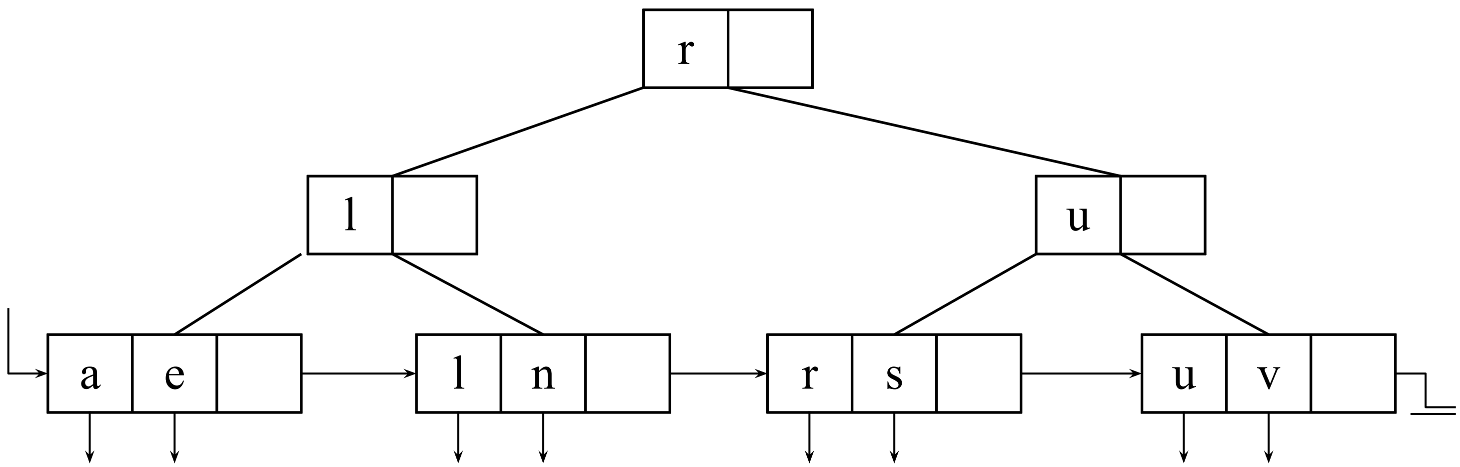
«j+1»

r s | u v

Overflow(B) □ Dividir en la mitad.

$$\underbrace{l \quad r \quad u}_{n = 3 \Rightarrow \left\lceil \frac{n}{2} \right\rceil = 2}$$

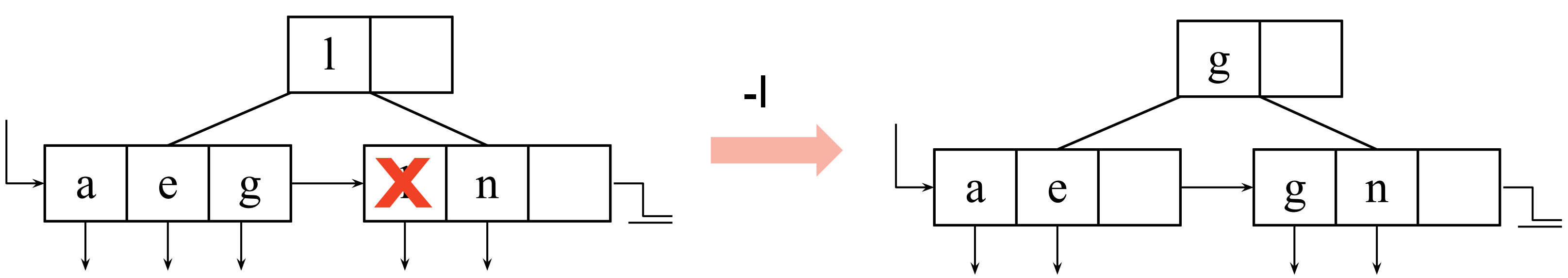
h) Familia de árboles B (B+)



h) Familia de árboles B (B+)

- **Eliminación**
  - ***SIEMPRE*** se elimina a nivel de hoja.
  - Si no hay *underflow*, solo se ajustan las claves de las hojas (los índices se conservan).
  - Si hay *underflow*, hacer:

    1. **Redistribución** *(si el hermano puede, le pasa uno)*
       Primero se ve el hermano derecho, luego el izquierdo.
       El dato pasa directo y luego se ajusta el padre de ser necesario.



h) Familia de árboles B (B+)

2. **Fusión** *(cuando no se puede redistribuir)*

h) Familia de árboles B (B+)

- Ejemplo: Eliminar en un árbol B+ con M=3

  I) Índice: | | |

$$Cant.\ máx\ hijos = M = 3 \Rightarrow Cant.\ máx\ claves = M - 1 = 2$$

$$Cant.\ min\ hijos = \left\lceil \frac{M}{2} \right\rceil = 2 \Rightarrow Cant.\ min\ claves = \left\lceil \frac{M}{2} \right\rceil - 1 = 1$$

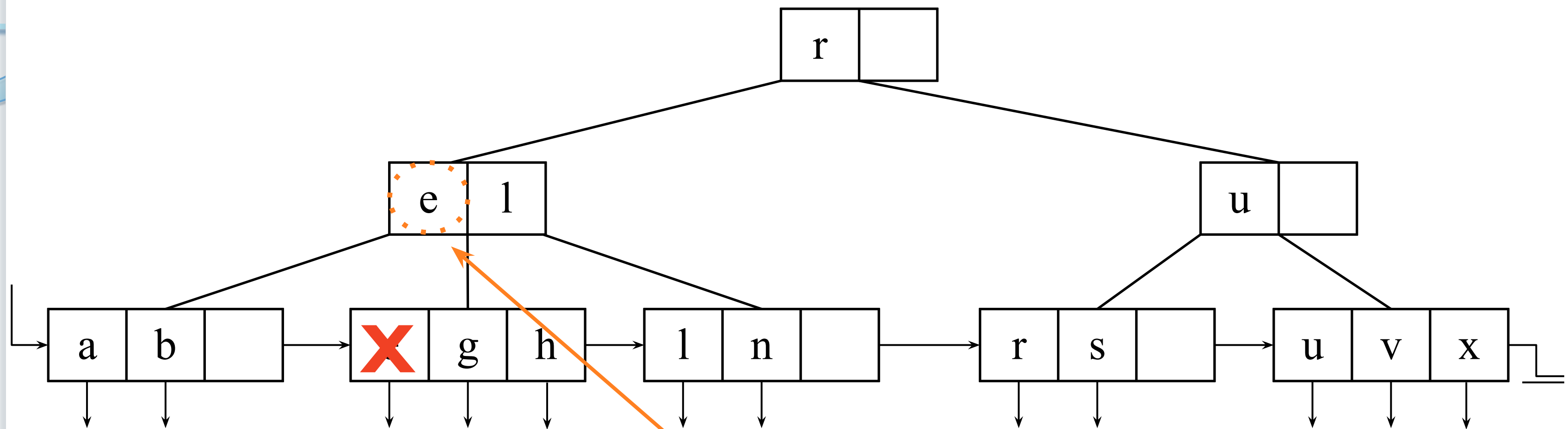
  II) Conjunto secuencia: | | | |

$$Cant.\ máx\ claves = M_{HOJA} = 3$$

$$Cant.\ min\ claves = \left\lceil \frac{M_{HOJA}}{2} \right\rceil = 2$$

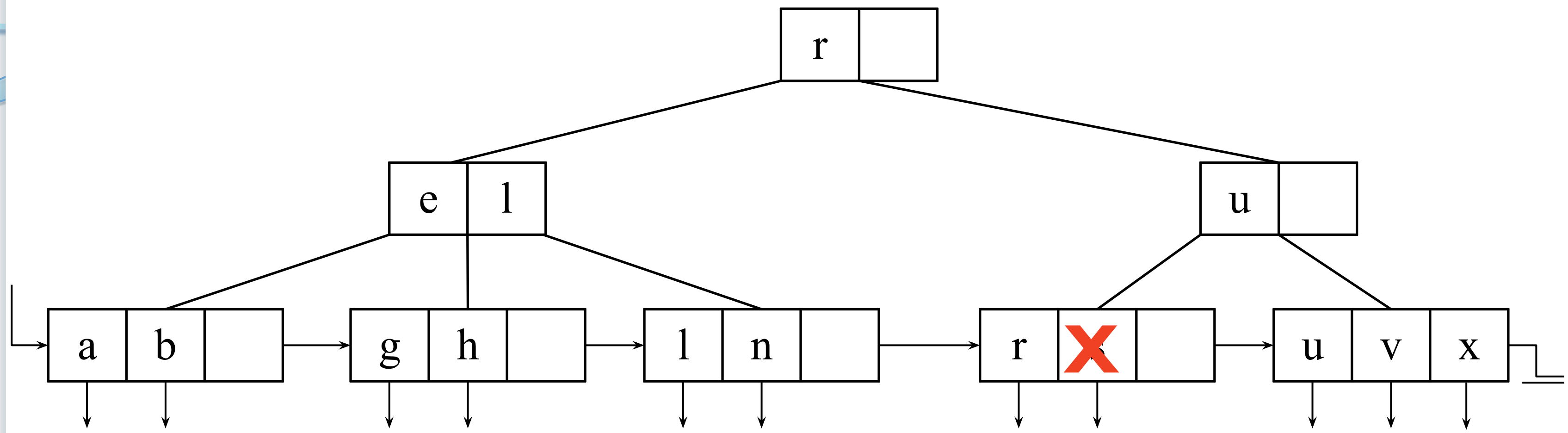
$$j = \left\lceil \frac{M_{HOJA} + 1}{2} \right\rceil = \left\lceil \frac{3 + 1}{2} \right\rceil = 2$$

h) Familia de árboles B (B+)



-e Se elimina de la hoja.
No es necesario eliminarla del índice.

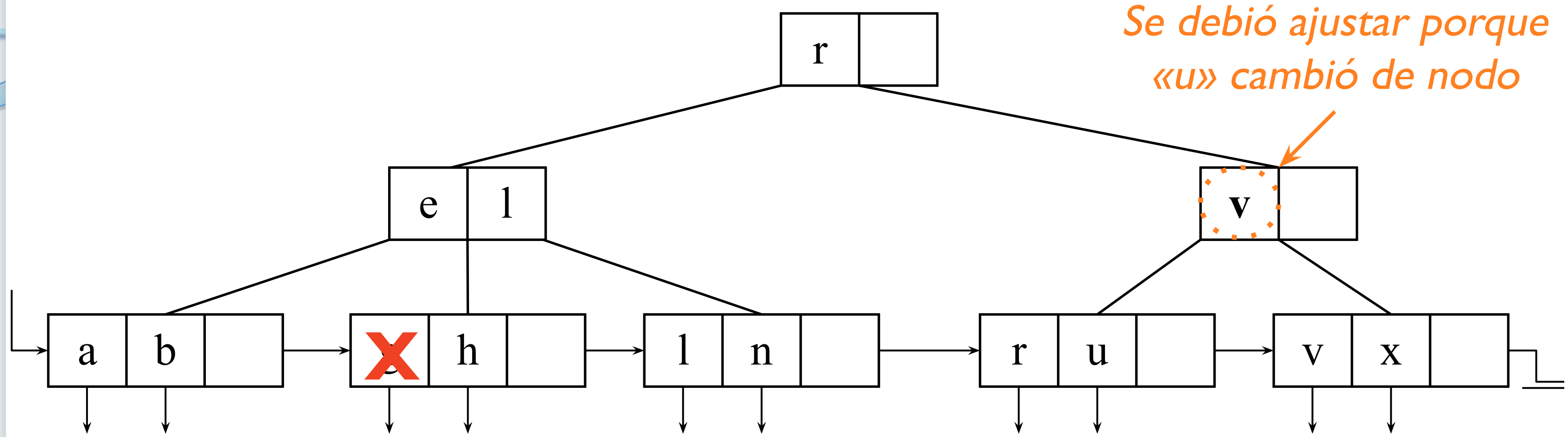
h) Familia de árboles B (B+)



-s *Underflow* □ 1° ¿Redistribuir?

h) Familia de árboles B (B+)

Se debió ajustar porque «u» cambió de nodo



-g **Underflow** □ 1° ¿Redistribuir? ✗

2° Fusionar: 2 hermanos

$$n = 5 \Rightarrow \left\lceil \frac{n}{2} \right\rceil = 3$$

$\underbrace{a \quad b}_{\textit{Hermano Izq}} \quad \underbrace{h}_{\textit{Nodo Prob}} \mid \underbrace{l \quad n}_{\textit{Hermano Der}}$

h) Familia de árboles B (B+)