

**Συστήματα Διαχείρισης και Ανάλυσης Δεδομένων**  
**Διδάσκων: Ιωάννης Κωτίδης**

Εαρινό εξάμηνο 2023-2024

**Πρώτη Σειρά Ασκήσεων**

Ανάθεση: 22-03-2024

Παράδοση: 01-04-2024 Ώρα (23:55)

**Λύσεις Πρώτης Σειράς Ασκήσεων**

**Λύση Άσκησης 1**

1. Μέγεθος τομέα σε KB (kilobytes).

$$\text{Μέγεθος Τομέα} = 80\text{GB} / (4 * 2 * 512 * 1024) = 80\text{GB} / 4194304 = 83886080 \text{ KB} / 4194304 = 20 \text{ KB}$$

2. Αριθμό των κυλίνδρων που καταλαμβάνει η σχέση R.

$$\text{Η σχέση R καταλαμβάνει } 8192 * 40 \text{ KB} = 327680 \text{ KB} = 320 \text{ MB}$$

$$\text{Σε ένα ίχνος χωράνε } 1024 * 20\text{KB} = 20 \text{ MB.}$$

Για την αποθήκευση της R απαιτούνται  $320/20=16$  ίχνη δηλαδή  $16/8=2$  κύλινδροι.

3. Χρόνος που απαιτείται για την ανάγνωση ολόκληρης της σχέσης R

Σε μία περιστροφή του δίσκου η κεφαλή μπορεί να διαβάσει ένα ίχνος. Το αρχείο καταλαμβάνει 16 ίχνη. Συνεπώς ο χρόνος που απαιτείται είναι:

$$t = (60000 * 16) / 7200 = 133.33 \text{ ms.}$$

Στον χρόνο αυτό πρέπει να προσθέσουμε τον μέσω χρόνο μετακίνησης της κεφαλής καθώς και τον την μέση καθυστέρηση περιστροφής. Συνεπώς ο συνολικό χρόνος για το διάβασμα της σχέσης είναι:

$$8\text{ms} + (60000/7200) / 2 + 133.33 = 8\text{ms} + 4,16\text{ms} + 133.33 = 145,49 \text{ ms.}$$

4. Το ελάχιστο δυνατό μέγεθος ενός μπλοκ είναι 40 KB δηλαδή όσο απαιτείται για να χωρέσει μία εγγραφή της σχέσης R. Παρατηρούμε ότι ένα μπλοκ του δίσκου καταλαμβάνει 2 τομείς. Από το ερώτημα 3 έχουμε σε 133.33 ms η κεφαλή διαβάζει ένα ίχνος δηλαδή 1024 τομείς. Αρα διαβάζει δύο τομείς σε χρόνο  $2 * 133,33 \text{ ms} / 1024 = 0,26 \text{ ms}$ . Οπότε ο συνολικός χρόνος για την ανάγνωση 100 εγγραφών είναι:

$$100 * (8\text{ms} + 4,16\text{ms} + 0,26\text{ms}) = 1242\text{ms.}$$

## Λύση Άσκησης 2

1.

Μέγεθος εγγραφής δεδομένων =  $9+30+10+9+8+1+40+4+4=115$  bytes.

Μέγεθος εγγραφής ευρετηρίου πρώτου επιπέδου =  $AΦΜ + 6 = 9 + 6 = 15$  bytes

Σε ένα μπλοκ χωράνε  $\text{floor}(1024/115)=8$  εγγραφές του αρχείου, άρα ο αριθμός των block του πρώτου επιπέδου είναι  $\text{ceiling}(60000/8)=7500$  σελίδες.

2.

Σε ένα μπλοκ χωράνε  $1024/15=68$  εγγραφές του ευρετηρίου πρώτου επιπέδου. Συνεπώς απαιτούνται  $\text{ceiling}(7500/68)=111$  μπλοκ.

3.

Για το πρώτο επίπεδο απαιτούνται 111 μπλοκ, για το δεύτερο επίπεδο απαιτούνται  $\text{ceiling}(111/68)=2$  μπλοκ και για το τρίτο επίπεδο απαιτείται άλλο ένα μπλοκ.

4. Σύνολο μπλοκ ευρετηρίου  $111+2+1=114$  μπλοκ

5. Αριθμός προσπελάσεων =  $4(3+1)$ .

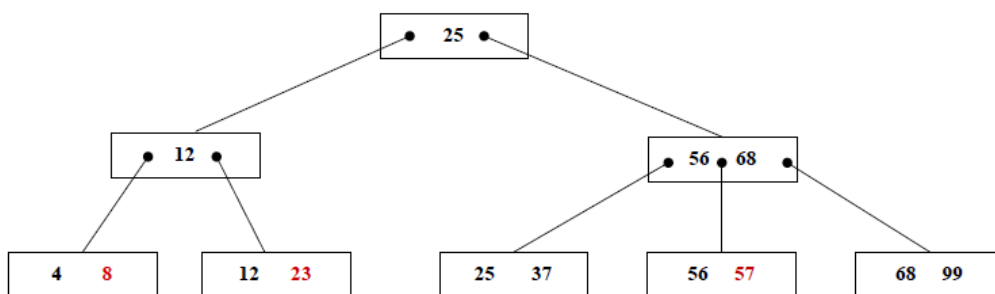
## Λύση Άσκησης 3

**Εισαγωγή τιμών: 8, 23, 57**

Η τιμή 8 είναι μικρότερη του 25, μικρότερη του 12 και μεγαλύτερη του 4. Συνεπώς θα τοποθετηθεί στον πρώτο από αριστερά κόμβο φύλο δεξιά της τιμής 4.

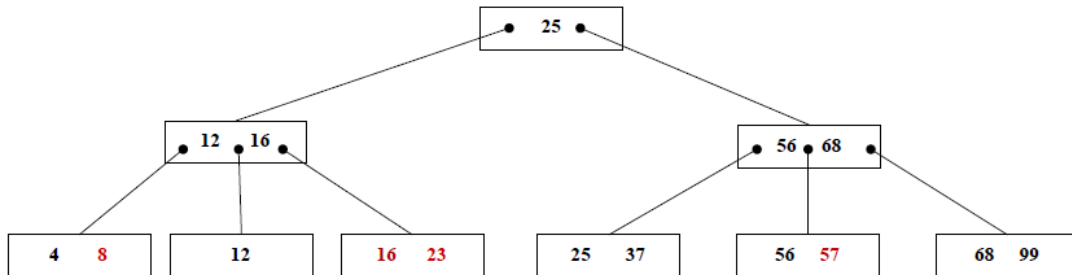
Η τιμή 23 είναι μικρότερη του 25, μεγαλύτερη του 12 και μεγαλύτερη του 13. Συνεπώς θα τοποθετηθεί στον δεύτερο από αριστερά κόμβο φύλο δεξιά της τιμής 12.

Η τιμή 57 είναι μεγαλύτερη του 25, μεγαλύτερη του 56 και μικρότερη του 68. Συνεπώς θα τοποθετηθεί στον τέταρτο από αριστερά κόμβο-φύλο δεξιά της τιμής 56.



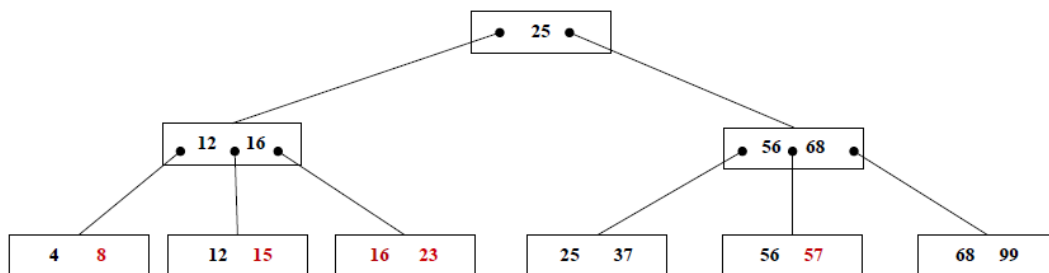
### Εισαγωγή τιμής 16

Η τιμή 16 είναι μικρότερη του 25, μεγαλύτερη του 12 και μικρότερη του 23. Συνεπώς θα τοποθετηθεί στον δεύτερο από αριστερά κόμβο ανάμεσα στις τιμές 12 και 23. Επειδή όμως η χωρητικότητα του κόμβου-φύλου είναι 2 ( $n=2$ ) η εισαγωγή της τιμής 16 θα προκαλέσει υπερχείλιση. Θα δημιουργηθεί ένας νέος κόμβος με τιμές **[16,23]**, ενώ ταυτόχρονα η τιμή 16 θα αντιγραφεί στον κόμβο γονέα με έναν δείκτη προς τον νέο κόμβο.



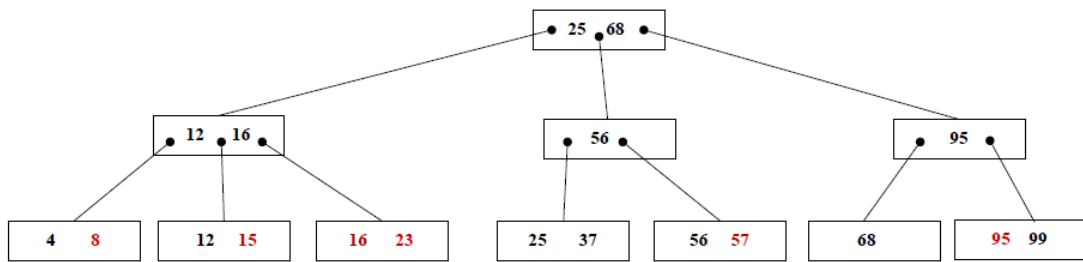
### Εισαγωγή τιμής 15

Η τιμή 15 είναι μικρότερη του 25, μεγαλύτερη του 12 και μικρότερη του 16. Συνεπώς θα τοποθετηθεί στον δεύτερο από αριστερά κόμβο-φύλο δεξιά της τιμής 12.



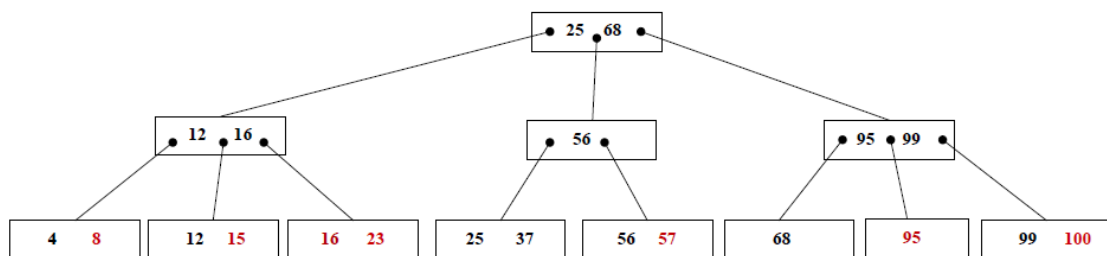
### Εισαγωγή τιμής 95

Η τιμή 95 είναι μεγαλύτερη του 25, μεγαλύτερη του 68 και μικρότερη του 99. Συνεπώς θα τοποθετηθεί στον τελευταίο δεξιό κόμβο φύλο ανάμεσα στις τιμές 68 και 99. Ο κόμβος θα υπερχείλισει οπότε και θα δημιουργηθεί ένας νέος κόμβος με τιμές **[95,99]**, ενώ ταυτόχρονα η τιμή 95 θα αντιγραφεί στον κόμβο γονέα. Η εισαγωγή της τιμής 95 στον κόμβο γονέα [56,68] θα προκαλέσει υπερχείλιση οπότε θα δημιουργηθεί ένας νέος κόμβος [68,95]. Παρατηρούμε ότι η τιμή 68 περισεύει, οπότε και ανεβαίνει στον κόμβο ρίζα.



### Εισαγωγή τιμής 100

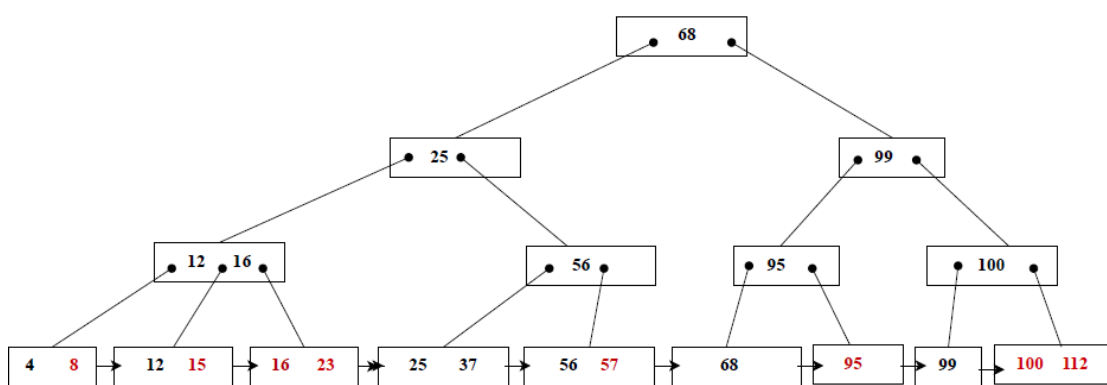
Η τιμή 100 είναι μεγαλύτερη του 68, μεγαλύτερη του 95 και μεγαλύτερη του 99. Συνεπώς θα τοποθετηθεί στον τελευταίο δεξιό κόμβο φύλο δεξιά της τιμής 99. Ο κόμβος θα υπερχειλίσει οπότε και θα δημιουργηθεί ένας νέος κόμβος με τιμές [99,100], ενώ ταυτόχρονα η τιμή 99 θα αντιγραφεί στον κόμβο γονέα.



### Εισαγωγή τιμής 112

Η τιμή 112 είναι μεγαλύτερη του 68, μεγαλύτερη του 99 και μεγαλύτερη του 100. Οπότε θα τοποθετηθεί στον τελευταίο δεξιό κόμβο φύλο μετά την τιμή 100. Ο κόμβος θα υπερχειλίσει οπότε θα δημιουργηθεί ένας νέος κόμβος-φύλο [100,112] και η τιμή 100 θα αντιγραφεί στον κόμβο γονέα.

Ο κόμβος γονέα θα υπερχειλίσει οπότε θα δημιουργηθεί ένας νέος κόμβος στο επίπεδο 1 [99,100] και η τιμή 99 θα αντιγραφεί στην ρίζα του δέντρου. Ο κόμβος ρίζα θα υπερχειλίσει με αποτέλεσμα να δημιουργηθεί ένας νέος κόμβος [68,99]. Παρατηρούμε ότι ο κόμβος 68 περισεύει οπότε και μετακινείται ένα επίπεδο επάνω και σχηματίζει την ρίζα του δέντρου.



### Αναζήτηση: $A \geq 12$ AND $A \leq 37$

Θα προσπελάσουμε τους εξής κόμβους:

$[68] \rightarrow [25] \rightarrow [12,16] \rightarrow [12,15] \rightarrow [16,23] \rightarrow [25,37]$

Επειδή κάθε κόμβος αποθηκεύεται σε ένα μπλοκ θα προσπελάσουμε 6 block του ευρετηρίου.

Σε αυτά θα πρέπει να προσθέσουμε και τα μπλοκ του αρχείου που περιέχει τις εγγραφές. Με βάση το δέντρο θα ανακτηθούν συνολικά 6 εγγραφές (εγγραφές με κλειδί: 12,15,16,23,25 και 37).

Συνολικός αριθμός μπλοκ που θα προσπαλεστούν =  $6 + 6 = 12$ .

### Λύση Άσκησης 4

1. Εισαγωγή 1000

1000	

**U=1/6**      17%

**0**      **1**

2. Εισαγωγή 0000

0000	
1000	

**U=2/6**      33%

**0**      **1**

3. Εισαγωγή 1101

0000	
1000	1101

**U=3/6**      50%

**0**      **1**

4. Εισαγωγή 0010

0010	
0000	
1000	1101

**U=4/6**      67%

**0**      **1**

5. Εισαγωγή 0010

0010		
0010		
0000		
1000	1101	$U=5/9$ 56%
<b>0</b>	<b>1</b>	

6. Εισαγωγή 1100

1100		
0010		
0010		
0000		
1000	1101	$U=6/9$ 67%
<b>0</b>	<b>1</b>	

7. Εισαγωγή 0011

1100		
0010		
0010		
0000	0011	
1000	1101	$U=7/9$ 78%
<b>0</b>	<b>1</b>	

Utilization =78%, ι αυξάνεται σε 2 και το m=2(10).

1100		
0000	0011	0010
1000	1101	0010
<b>00</b>	<b>*1</b>	<b>10</b>

$U=7/9$  | 78%

Utilization =78%, οπότε m=3(11).

1100			
0000		0010	
1000	1101	0010	0011
<b>00</b>	<b>01</b>	<b>10</b>	<b>11</b>

$U=7/12$  | 58%

8. Εισαγωγή 1111

1100				
0000		0010	1111	
1000	1101	0010	0011	
<b>00</b>	<b>01</b>	<b>10</b>	<b>11</b>	

U=8/12 | **67%**

9. Εισαγωγή 0110

1100		0110		
0000		0010	1111	
1000	1101	0010	0011	
<b>00</b>	<b>01</b>	<b>10</b>	<b>11</b>	

U=9/12 | **75%**

Utilization =75%, ι αυξάνεται σε 3 και το m=4(100).

		0110		
0000		0010	1111	
1000	1101	0010	0011	1100
<b>000</b>	<b>*01</b>	<b>*10</b>	<b>*11</b>	<b>100</b>

U=9/15 | **60%**

10. Εισαγωγή 1110

		1110		
		0110		
0000		0010	1111	
1000	1101	0010	0011	1100
<b>000</b>	<b>*01</b>	<b>*10</b>	<b>*11</b>	<b>100</b>

U=10/18 | **56%**

### Λύση Άσκησης 5

1) Για τις επερωτήσεις τύπου Q1 πρέπει να λάβουμε υπόψη μας τα  $10-N$  bits από την εφαρμογή της συνάρτησης  $h_2$ . Άρα θα προσπελαστούν:  $2^{(10-N)}$  μπλοκ. Αντιθέτως για τις επερωτήσεις τύπου Q2 λαμβάνουμε υπόψη μας τα  $N$  bits από τη συνάρτηση  $h_1$ , άρα θα προσπελαστούν  $2^N$  μπλοκ.

2) Με δεδομένο ότι το 30% των επερωτήσεων είναι τύπου Q1 και το 70% τύπου Q2, ο μέσος αριθμός των μπλοκ που πρέπει να προσπελαστούν για την απάντηση των επερωτήσεων στην R δίνεται από τον τύπο:  $0.3 \cdot 2^{(10-N)} + 0.7 \cdot 2^N$