

Συστήματα Διαχείρισης και Ανάλυσης Δεδομένων
Διδάσκων: Ιωάννης Κωτίδης

Εαρινό εξάμηνο 2024-2025

Λύσεις Πρώτη Σειρά Ασκήσεων

Λύση Άσκησης 1

a)

Συνολικός Αριθμός τομέων = 4 πλακέτες * 2 επιφάνειες * 4096 ίχνη * 1024 τομείς

Συνολικός Αριθμός τομέων = $4 * 2 * 4096 * 1024 = 33.554.432$

Ένα μπλοκ είναι 4096 bytes δηλαδή $4096/512 = 8$ τομείς. Συνεπώς ο συνολικός αριθμός των μπλοκ είναι: $33.554.432/8 = 4.194.304$

Κάθε μπλοκ χωράει $4096/256 = 16$ εγγραφές. Συνεπώς,

Συνολικός αριθμός εγγραφών = $4.194.304 * 16 = 67.108.864$

b)

Συνολικά μπλοκ αρχείου = $1.000.000 * 256/4096 = 62500$

Σε μία περιστροφή ο δίσκος διαβάζει 1024 τομείς δηλαδή $1024/8 = 128$ μπλοκ. Δηλαδή διαβάζει 128 μπλοκ σε $(1/7200) * 60 = 0,00833$ sec = **8,33 ms**.

Συνεπώς διαβάζει το 1 μπλοκ σε $0,00833/128 = 0,000065$ sec = **0,065 ms**.

Χρόνος Μεταφοράς Μπλοκ (tb) = **0,065 ms**

Χρόνος μισής περιστροφής (trd) = $0,00833/2 = 0,004165$ sec = **4,17 ms**

Μέσος χρόνος μετακίνησης κεφαλής (ts) = **10 ms**

Χρόνος ανάγνωσης ολόκληρου του αρχείου = $ts + trd + tb * (62500) =$
 $10 \text{ ms} + 4,17 \text{ ms} + (0,065 * 62500) = 4076,67 \text{ ms} = 4,07667 \text{ sec}$

Έστω X ο αριθμός των εγγραφών που ο χρόνος για τυχαία προσπέλαση είναι μεγαλύτερος από τον απαιτούμενο χρόνο για πλήρη ανάγνωση του αρχείου:

$$\begin{aligned} X(ts + trd + tb) &> 4076,67 \text{ ms} \\ X(10 + 4,17 + 0,065) &> 4076,67 \\ X &> 4076,67 / 14,235 = \\ X &> 286,38 \rightarrow \mathbf{X > 287} \end{aligned}$$

Συνεπώς αν θέλουμε να αναζητήσουμε περισσότερες από 287 εγγραφές μας συμφέρει να διαβάσουμε ολόκληρο το αρχείο.

Λύση Άσκησης 2

a.

Μέγεθος εγγραφής=6+25+25+8=64 bytes. Σε ένα μπλοκ χωράνε $\text{floor}(512/64)=8$ εγγραφές.
Πλήθος μπλοκ αρχείου = $100000/8=12500$ μπλοκ

b.

Κάθε εσωτερικός κόμβος του B+ δέντρου μπορεί να έχει n κλειδια και $n+1$ δείκτες και όλα αυτά πρέπει να χωράνε σε ένα μπλοκ. Συνεπώς ο **μέγιστος αριθμός κλειδιών** κάθε εσωτερικού κόμβου υπολογίζεται ως εξής:

$$(n*8) + (n+1)*4 \leq 512 \rightarrow 12n+4 \leq 512 \rightarrow 12n \leq 508 \rightarrow n \leq \text{floor}(508/12) \rightarrow n \leq \text{floor}(42,33) \rightarrow n \leq 42.$$

Συνεπώς σε έναν ενδιάμεσο κόμβο χωράνε **$n=42$ κλειδιά και $n+1=43$ δείκτες**. Με άλλο λόγια κάθε ενδιάμεσος κόμβος μπορεί να έχει 43 παιδιά.

Οι κόμβοι-φύλλα του B+ δένδρου θα έχουν το ίδιο πλήθος κλειδιών και δεικτών, εκτός του ότι οι δείκτες είναι δείκτες δεδομένων και υπάρχει ένας δείκτης προς το επόμενο μπλοκ. Συνεπώς έχουμε:

$$(n * (8+5)) + 4 \leq 512 \rightarrow 13n \leq 508 \rightarrow n \leq \text{floor}(508/13) \rightarrow n \leq \text{floor}(39,07) \leq 39, n=39$$

Από τα παραπάνω έχουμε:

- Μέγιστος αριθμός παιδιών ενδιάμεσου κόμβου = **43** και ελάχιστος αριθμός παιδιών= $\text{ceiling}(43/2)=22$.
- Μέγιστος αριθμός κλειδιών κόμβου φύλλου= **39** και ελάχιστος αριθμός κλειδιών = $\text{ceiling}(39/2)=20$

Το ευρητήριο καταλαμβάνει τα λιγότερα μπλοκ όταν οι κόμβοι (φύλλα και ενδιάμεσοι) είναι γεμάτοι. Συνεπώς χρειαζόμαστε:

Αριθμός μπλοκ πρώτου επιπέδου (φύλλων) $\text{ceiling}(100000/39)=2564,103=2565$

Αριθμός μπλοκ δεύτερου επιπέδου= $\text{ceiling}(2565/43)=60$

Αριθμός μπλοκ τρίτου επιπέδου= $\text{ceiling}(60/43)=2$

Αριθμός μπλοκ πρώτου επιπέδου= $\text{ceiling}(2/43)=1$

Ελάχιστος αριθμός μπλοκ του ευρητηρίου= $1+2+60+2565=2628$

Το ευρητήριο καταλαμβάνει τα περισσότερα μπλοκ όταν οι κόμβοι (φύλλα και ενδιάμεσοι) περιέχουν τον ελάχιστο αριθμό δεικτών/τιμών. Συνεπώς χρειαζόμαστε:

Αριθμός μπλοκ πρώτου επιπέδου (φύλλων) $\text{ceiling}(100000/20)=5000$

Αριθμός μπλοκ δεύτερου επιπέδου= $\text{ceiling}(5000/22)=228$

Αριθμός μπλοκ τρίτου επιπέδου= $\text{ceiling}(228/22)=11$

Αριθμός μπλοκ τέταρτου επιπέδου= $\text{ceiling}(11/22)=1$

Μέγιστος αριθμός μπλοκ του ευρητηρίου= $1+11+228+5000=5240$

c. Θα διαβάσουμε:

Την ρίζα του ευρετηρίου (**1 block**)

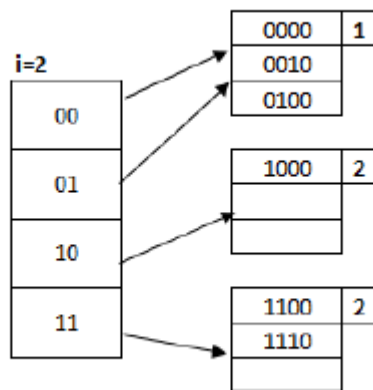
Ένα μπλοκ στο δεύτερο επίπεδο και ένα μπλοκ στο τρίτο επίπεδο ώστε να εντοπίσουμε το μπλοκ φύλλο που περιέχει την πρώτη τιμή μεγαλύτερη του d1. (**2 block**)

Στη συνέχεια θα διαβάσουμε τους κόμβους φύλλα με τα 1000 κλειδιά. Από το πρώτο ερώτημα γνωρίζουμε ότι ένας κόμβος φύλλο περιέχει 39 κλειδιά, οπότε $\text{ceiling}(1000/39) = 26$ μπλοκ

Τέλος θα διαβάσουμε **1000** μπλοκ για να ανακτήσουμε τις 1000 εγγραφές.

Σύνολο = $1+2+26+1000=1029$ μπλοκ.

Λύση Άσκησης 3



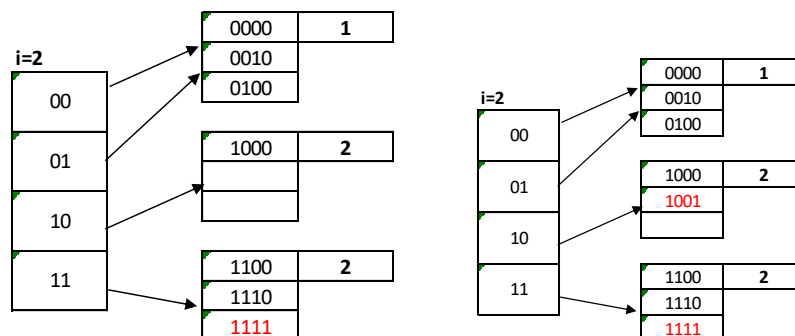
a.

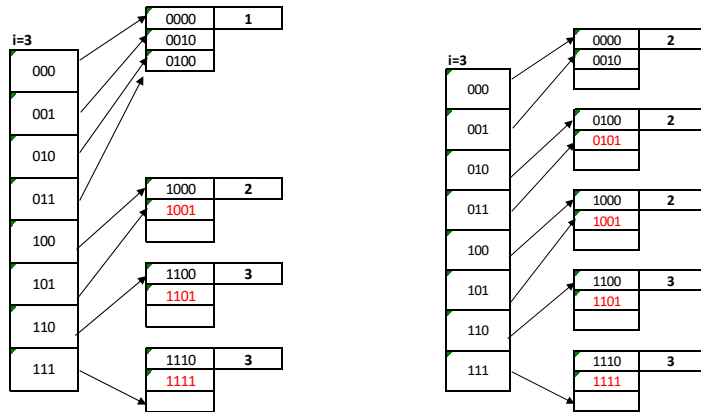
Έξι κλειδιά: Έχουμε δύο κάδους με τοπικό βάθος 2 στους οποίους μπορούμε να καταχωρήσουμε 3 κλειδιά. Επιπλέον υπάρχει ένας κάδος με τοπικό βάθος 1 τον οποίο αν διασπάσουμε μπορούμε να καταχωρήσουμε άλλα τρία κλειδιά.

b.

Δύο κλειδιά. Η καταχώρηση δύο κλειδιών στον τρίτο κάδο (κάδος με τιμές 1100, 1110) θα οδηγήσει σε αύξηση του ολικού βάθους του ευρετηρίου.

c.



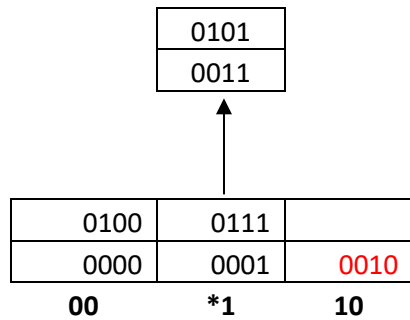


Λύση Άσκησης 4

a. $i=2, m=10$ (2)

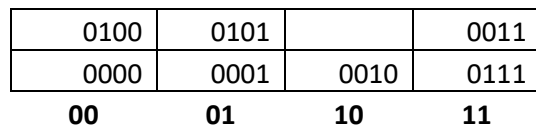
b.

1. Εισαγωγή **0010**



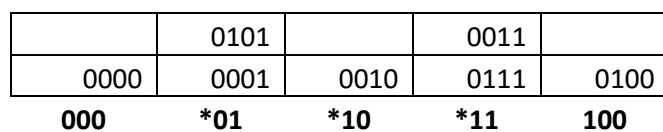
Utilization **87,5%**

Utilization 87,5% $\geq 80\%$, και το m γίνεται **11** (3)



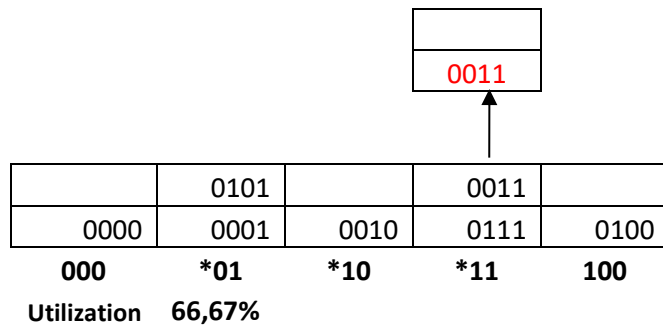
Utilization **87,5%**

Utilization 87,5% $\geq 80\%$, το m γίνεται **100** (4) και $i=3$

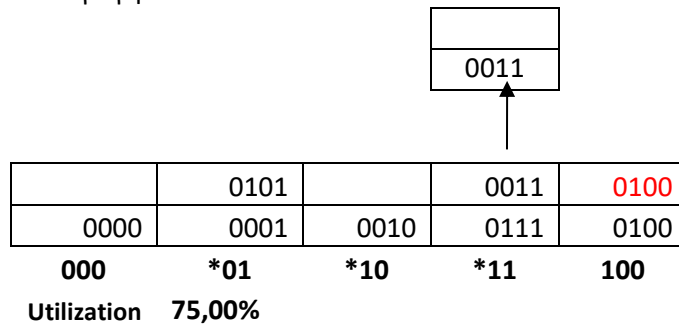


Utilization **70%**

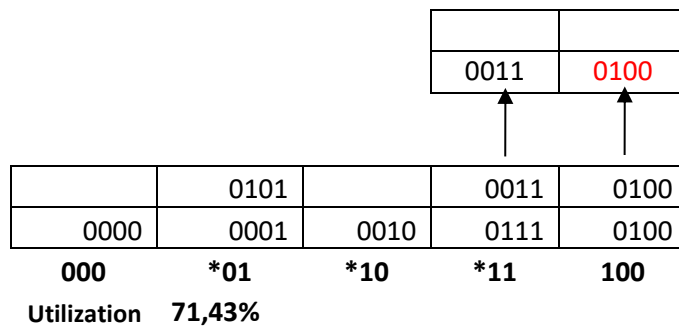
2. Εισαγωγή 0011



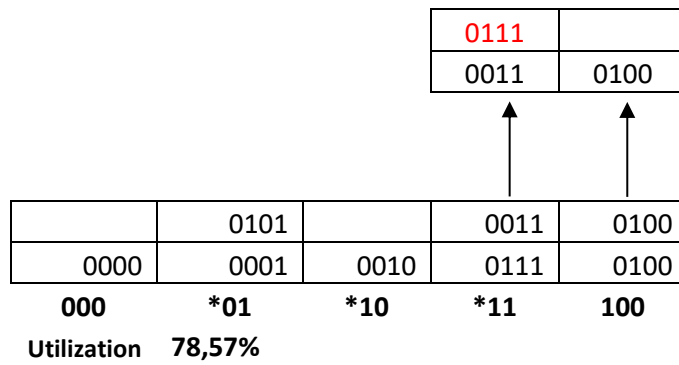
3. Εισαγωγή 0100



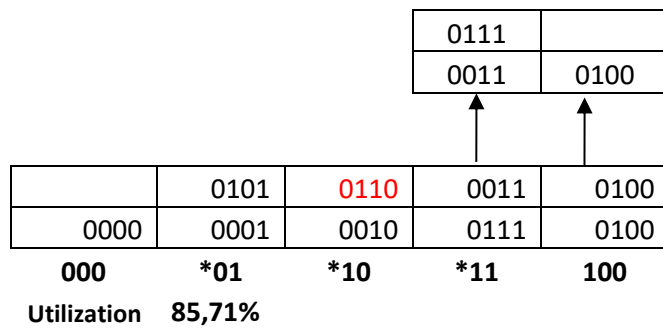
4. Εισαγωγή 0100



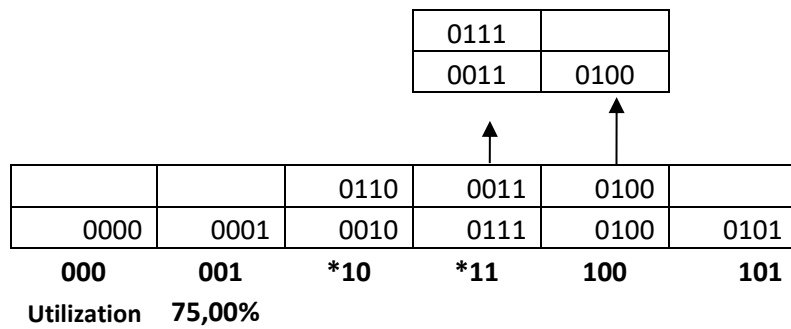
5. Εισαγωγή 0111



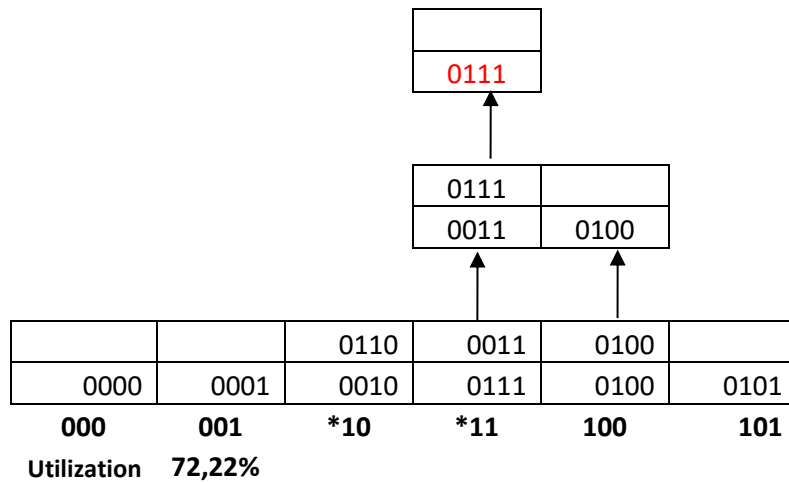
6. Εισαγωγή 0110



Utilization 87,5% >=80%, η αυξάνει σε 101 (5).



7. Εισαγωγή 0111



c.

Ο μέσος αριθμός προσπελάσεων είναι:

- 9 εγγραφές απαιτούν 1 προσπέλαση
- 3 εγγραφές απαιτούν 2 προσπελάσεις
- 1 εγγραφή απαιτεί 3 προσπελάσεις

Συνεπώς έχουμε: $(9 * (1/13)) + (3 * (2/13)) + (1 * (3/13)) = 1,384$

Επιπλέον ερώτημα

Το συγκεκριμένο ερώτημα **ΔΕΝ** αποτελεί μέρος της λύσης τα άσκησης. Παρατίθεται διότι συζητήθηκε στο φροντιστήριο. Το ερώτημα είναι: «Ποιος είναι ο μέσος αριθμός των προσπελάσεων όταν η τιμή του κλειδιού που αναζητούμε ΔΕΝ υπάρχει στο ευρετήριο;».

Μέσος αριθμός Προσπελάσεων όταν το κλειδί δεν υπάρχει

Αν αναζητούμε τιμές που δεν υπάρχουν τότε εξετάζοντας τα 3 τελευταία bits το κόστος αναζήτησης θα είναι:

- 000 → 1
- 001 → 1
- 010 → 1
- 011 → 3
- 100 → 2
- 101 → 1
- 110 → 1
- 111 → 3

Συνεπώς έχουμε: $(5 * (1/8)) + (1 * (2/8)) + (2 * (3/8)) = 13/8 = 1,625$

Λύση Άσκησης 5

a. $150 \leq \text{salary} < 300$ and $\text{age} < 47$

b. $80 \leq \text{salary} < 150$ and $\text{age} < 60$