Week 13 Accessing and Indexing Data (1)

Assignment Solutions

1. For a disk of seek time = 5 ms, rotational latency = 3 ms, and transfer time = 0.0001 ms per kb, what's the cost of sequential access to get the next 100 kb of data?

- ○ 8 ms
- ○ 8.0001 ms
- ○ 0.01 ms

Solution: C

Explanation: Sequential access only costs transfer time, which is 0.0001ms/kb * 100kb = 0.01ms.

2. Consider an ISAM with fanout $F = 341$ and height $h = 3$. What's the maximum number of tuples it can store?

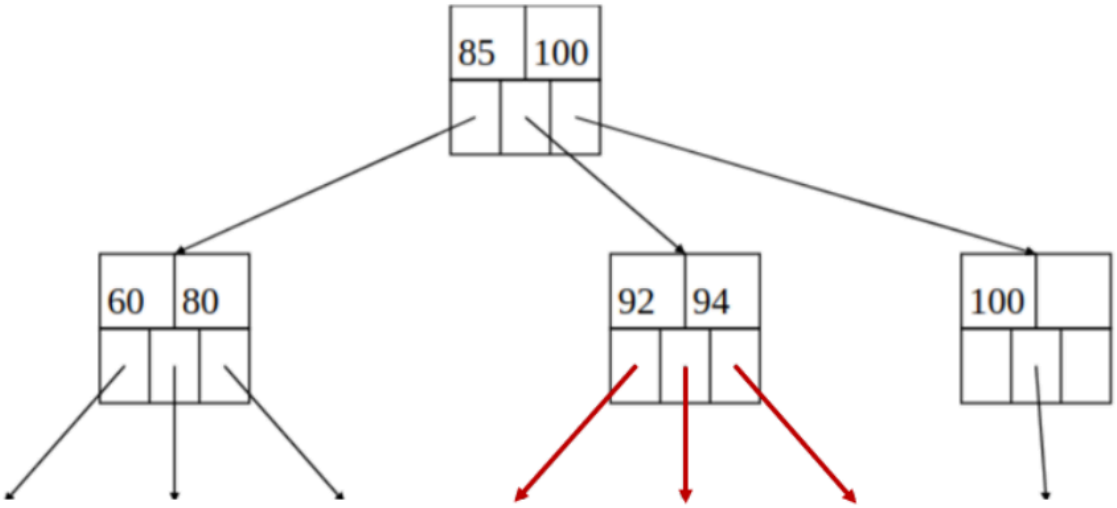
Note: The height of an index tree is the number of pointer hops from its root to reach the data tuples.

- ○ 116281
- ○ 1000000
- ○ 39651821

Solution: C

Explanation: The total number of tuples available is F^h, where F is the fanout and h is the height, because a node at each of the h levels can point to F child nodes. Therefore, $F^h = 341^3 = 39651821$

3. Considering the following ISAM, what range of values do the pointers in red represent?

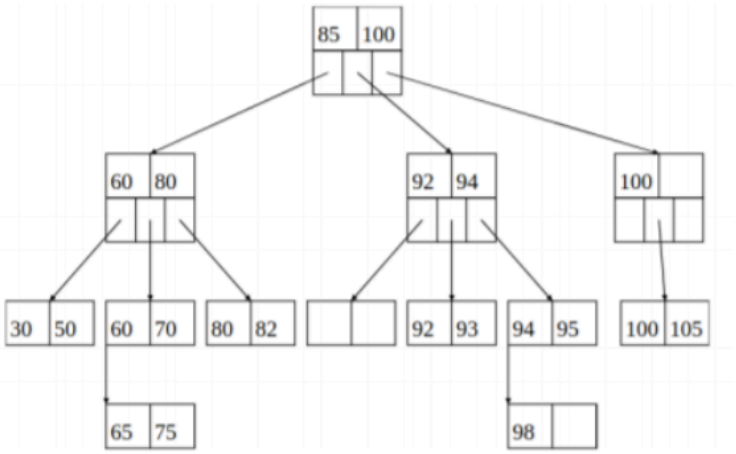


- ○ (85, 92], (92, 94], (94,100]
- ○ [-∞, 92), [92,94), [94, ∞)
- ○ [85, 92), [92, 94), [94, 100)

Solution: C

Explanation: From the root node, we can identify the range of the middle child node to be [85, 100). Thus, the ranges of the three red pointers are further divisions of [85, 100), which becomes [85, 92), [92, 94), and [94, 100).

4. What's the cost of looking up 65 in the following ISAM? Here, we define "cost" as the **number of random accesses** (disk pointers) to reach data tuples.



- 3
- 2
- 5
- 4

Solution: D

Explanation: To reach the data tuple of key 65, we need to go through 3 pointers to reach 65 in ISAM and then another pointer to locate the data tuple.

5. Compared to table scan and binary search, what are the advantages of ISAM?

- [ ] ISAM takes less time to complete a query.
- [ ] ISAM automatically sort overflowed pages to balance look up time for any keys.
- [ ] ISAM requires fewer disk random accesses.

Solution: AC

Explanation: Table scan takes O(n), and binary search takes $O(\log_2 n)$. ISAM takes $O(\log_F n)$, where F is the fanout. As fanout is usually much higher than 2, ISAM requires fewer disk pointer traversals (random accesses) -- and thus also takes less time to complete a query.