第十一章作业答案

XX YYY

Email: xx@scut.edu.cn

华南理工大学软件学院

May 19, 2023

Current Section

21.Why is a hash structure not the best choice for a search key on which range queries are likely?

参考答案:

A range query cannot be answered efficiently using a hash index, we will have to read all the buckets. This is because key values in the range do not occupy consecutive locations in the buckets, they are distributed uniformly and randomly throughout all the buckets.

Current Section

22.Suppose there is a relation r(A, B,C), with a B+-tree index with search key (A, B).

a. What is the worst-case cost of finding records satisfying $10 < A < 50$ using this index, in terms of the number of records retrieved n1 and the height h of the tree?

b. What is the worst-case cost of finding records satisfying $10 < A < 50 \bigwedge 5 < B < 10$ using this index, in terms of the number of records n2 that satisfy this selection, as well as n1 and h defined above?

c. Under what conditions on n1 and n2 would the index be an efficient way of finding records satisfying $10 < A < 50 \bigwedge 5 < B < 10$?

- 假设属性A的取值范围0 60，B的取值范围4 15，那么本B+树叶子层可能的布局是：

$$\ldots, \underbrace{(10,4),(10,5),\cdots,(10,15)}_{12}, \cdots, \underbrace{(11,4),\cdots,(11,15)}_{12}, \cdots, \underbrace{(60,4),\cdots,(60,15)}_{12}$$

- 查找过程与普通的B+树没有区别，
- 1. 从根走到叶子层，在叶子层找到目标记录地址
- 2. 然后访问目标记录

最坏的情况，就是目标记录的在磁盘上的分布不是连续的（索引是辅助索引），有多少条满足条件的记录，需要访问磁盘多少次。

引入参数:

- c1: 第1题, 叶子层满足要求的指针占据的磁盘块数
- c2: 第2题,叶子层满足要求的指针占据的磁盘块数

1. 访问磁盘的次数:

$$h + n1 + c1$$

每次访问都要寻道+传输

2. 访问磁盘的次数:

$$h + n2 + c2$$

每次访问都要寻道+传输，注意$n1 > n2$

3. $n1 = n2 = 1$ 这是最优情况