

Lab-7 Exercises

You are given the following information:

Executives is a relation and has the following attributes:
ename, *title*, *dname*, and *address*; all are string fields of the same length.

The *ename* attribute is a candidate key.
The relation contains 10,000 pages.

1.

Consider the following query:

```
SELECT E.title, E.ename FROM Executives E WHERE E.title = 'CFO'
```

Assume that only 10% of Executives tuples meet the selection condition.

- (a) Suppose that a clustered B+ tree index on title is (the only index) available. What is the cost of the best plan? (In this and subsequent questions, be sure to describe the plan you have in mind.)
- (b) Suppose that an unclustered B+ tree index on title is (the only index) available. What is the cost of the best plan?
- (c) Suppose that a clustered B+ tree index on ename is (the only index) available. What is the cost of the best plan?
- (d) Suppose that a clustered B+ tree index on address is (the only index) available. What is the cost of the best plan?

2.

Suppose that the query is as follows:

```
SELECT E.ename FROM Executives E WHERE E.title = 'CFO' AND E.dname = 'Toy'
```

Assume that only 10% of Executives tuples meet the condition $E.title = 'CFO'$, only 10% meet $E.dname = 'Toy'$, and that only 5% meet both conditions.

- (a) Suppose that a clustered B+ tree index on title is (the only index) available. What is the cost of the best plan?

- (b) Suppose that a clustered B+ tree index on dname is (the only index) available. What is the cost of the best plan?
- (c) Suppose that a clustered B+ tree index on <title, dname> is (the only index) available. What is the cost of the best plan?
- (d) Suppose that a clustered B+ tree index on <title, ename> is (the only index) available. What is the cost of the best plan?

3.

Suppose that the query is as follows:

```
SELECT E.title, COUNT(*) FROM Executives E GROUP BY E.title
```

- (a) Suppose that a clustered B+ tree index on title is (the only index) available. What is the cost of the best plan?
- (b) Suppose that an unclustered B+ tree index on title is (the only index) available. What is the cost of the best plan?
- (c) Suppose that a clustered B+ tree index on <title, ename> is (the only index) available. What is the cost of the best plan?

Solution

1.

(a) The best plan, a B+ tree search, would involve using the B+ tree to find the first title index such that title='CFO', cost = 2. Then, due to the clustering of the index, the relation pages can be scanned from that index's reference.

$$cost = 10000 * 10\% + 2500 * 10\% (\text{Scanning the index}) = 1000 + 250 + 2 = 1252 (\text{total cost}).$$

(b)

$$cost = 10000 * 10\% + 2500 * 10\% * (2+1)$$

(c)

Due to the WHERE clause, the clustered B+ index on ename doesn't help at all. The best alternative is to use a filescan, cost = 10000.

(d) Again, as in the previous answer, the best choice is a filescan, cost = 10000.

2.

(a)

A clustered index on title would allow scanning of only the 10% of the desired tuples. Thus, the total cost is $2(\textit{lookup}) + 10000 * 10\% + 2500 * 10\% = 1252$.

(b)

A clustered index on dname works functionally in the same manner as that in the previous question, for a cost $1002 + 250 = 1252$. The ename field still must be retrieved from the relation data pages.

(c)

In this case, using the index lowers the cost of the query slightly, due to the greater selectivity of the combined query and to the search key taking advantage of it. The total cost $= 2(\textit{lookup}) + 10000 * 5\% + 5000 * 5\% = 752$.

(d)

Although this index does contain the output field, the dname still must be retrieved from the relational data pages, for a cost of $2(\textit{lookup}) + 10000 * 5\% + 5000 * 10\% = 552$.

3.

(a)

Since title is the only attribute required, an index-only scan could be performed, with a running counter. This would cost $10000 * .25(\textit{index - only scan, smaller tuples}) + \textit{initial lookup in the index} = 2500 + 2 = 2502$.

(b)

Again, as the index contains the only attribute of import, an index-only scan could again be performed, for a cost of $2500 + 2 = 2502$.

(c)

The clustered B+ index given contains all the information required to perform an index-only scan, at a cost of $10000 * .5(\textit{tuple size}) = 5000 + 2 = 5002$.