CSE 344 Section 4: Relational Algebra, Indexing, Cost Estimation

Solution

Problem 1

Consider the following database schema:

Users(uid, name)
Comment(uid, pid, score, txt)
Picture(pid, uid, img)

(a)

Write a Relational Algebra Plan for the SQL query:

SELECT DISTINCT U.uid
FROM Users U, Picture P, Comment C
WHERE U.uid = P.uid
AND P.pid = C.pid
AND C.score > 8
GROUP BY U.uid, P.pid
HAVING COUNT(*) > 10

(b)

Write a Relational Algebra Plan for the SQL query:

SELECT P.pid
FROM Picture P
WHERE NOT EXISTS
(SELECT *
FROM Comment C
WHERE P.pid = C.pid
AND C.score < 5)
Problem 2

Consider a workload consisting of many queries of this kind:

```sql
SELECT x.name
FROM Member x, Tagged y, Picture z
WHERE x.mid = y.mid
AND y.pid = z.pid
AND z.year = ?;
```

For each index below, indicate if it can potentially speed up the workload, if it is the only index that exists. You will assume that the Member and Tagged are very large relations, and that only a very small number of pictures are in any given year.

Think of how the query will be executed and WHICH attribute will each table be accessed.

i. (1 point) Index on Picture(year).  
   Yes or no?

ii. (1 point) Index on Picture(pid).  
   Yes or no?

iii. (1 point) Index on Picture(pid, year).  
   Yes or no?

iv. (1 point) Index on Picture(year, pid).  
   Yes or no?

v. (1 point) Index on Tagged(mid).  
   Yes or no?

vi. (1 point) Index on Tagged(pid).  
   Yes or no?

vii. (1 point) Index on Tagged(pid, mid).  
   Yes or no?

viii. (1 point) Index on Member(mid).  
   Yes or no?

ix. (1 point) Index on Member(name).  
   Yes or no?

x. (1 point) Index on Member(age).  
   Yes or no?