Video Activities- Week-9

Submitted by: - Parth Parashar

Video 15.1: -

select \* from authors A, books B where A.id = B.authorid;

Text

Description automatically generated

For finding out the cost and hash value, we can run the select query as given below: -

Text

Description automatically generated

The steps followed for building a hash table are: -

1. Take the join attribute
2. Hash value = hash function(join attribute)
3. Insert the tuple into hash table based on the hash value.

Hashing works for the equality comparisons.

Simple case: - Cost: M+ N

Harder case: - Cost: 3(M+N)

Video 15.2: -

Join Activity

Select \* from authors A, books B where A.id = B.authorid; (index on Authors.id)

In this query, I will use the INL or Hash or Sort Merge. For hash, Hash table on authors.

Select \* from authors A, books B where A.id = B.authorid; (No index)

For this query, I will use Hash or Sort Merge. For hash, Hash Table on authors.

Block nested loops is super efficient and works with non-equi joins as well. Hash Join is efficient if one relation fits in memory and works only with equi-join.

Video 15.3: -

Aggregate Hash

Select count(\*) from musicians group by instruments;

Text

Description automatically generated

Result is: - (instrument, count)

(drums, 1)

(guitar, 1)

(vocals, 2)

Activity 16.1: -

Q-1: - select name, phone from students where sid = 1236;

Q-2: - select phone from students where name = ‘Mike’ and sid < 200;

Q-3: - select name , cid from students S JOIN isTaking I on S.sid = I.sid where name = ‘Xayne’;

Diagram, schematic

Description automatically generated

Diagram

Description automatically generated

Activity 16.2: -

1. How many rows does this query return?
2. 1 or 0
3. Assuming sid is the primary key or sid is unique.
4. Look at constraints.
5. Or look at # distinct values in sid column compared to the # of tuples
6. Select \* from students where sid < 1236;
7. Probably a lot
8. Need to know how many sids exist / # of students.
9. Are they sequentially assigned or removed
10. Things in catalogue that might help

* Min and max sid
* # distinct values in sid column
* Histogram of the distribution of the sid values