CSE 512: Distributed Database Systems

Project: Distributed Database System for a Smart Building

Part – 3

Query Optimization using Indexing in PostgreSQL:

Indexing is a database optimization technique that involves creating data structures to improve the speed of data retrieval operations on a database table. It is primarily aimed at enhancing the speed of data retrieval operations, such as SELECT queries. Without indexes, a database management system (DBMS) may need to perform a full table scan, checking every row to find the desired data. This can be inefficient for large datasets.

By default, PostgreSQL performs B-Tree indexing. They are the most used indexing technique. They are well-suited for various data types and provide efficient search, insertion, and deletion operations.

Well-designed indexes can significantly improve the overall performance of database applications. To showcase the impact of an index, we run the query to find all buildings containing at least one floor with more than four rooms with and without indexes on the buildings, floors, and rooms tables.

```
SELECT DISTINCT b.building_id, b.building_name

FROM buildings b

JOIN floors f ON b.building_id = f.building_id

JOIN rooms r ON f.floor_id = r.floor_id

GROUP BY b.building_id, b.building_name

HAVING COUNT(DISTINCT f.floor_id) > 0 AND COUNT(DISTINCT r.room_id) > 4;
```

We indexed buildings table on building_id, floors table on (floor_id, building_id), and rooms table on (room_id, floor_id)

The screenshot below is a clear indication of the improved query processing by PostgreSQL.