



COMP9120 Relational Database Systems

Tutorial Week 11: Storage and Indexing

Exercise 1. Calculating space and time

$$4 \times 1024 = 4096$$

Suppose we have a table Rel1(A, B, C). Each A field occupies 4 bytes, each B field occupies 12 bytes, each C field occupies 8 bytes. Rel1 contains 100,000 records. There are 100 different values for A represented in the database, 1000 different values for B, and 50,000 different values for C. Rel1 is stored with a primary B+ tree index on the composite key consisting of the pair of attributes (A,B); assume this index has 2 levels, including the root page and excluding the leaf (record) pages. Assume that in this database, each page is 4K bytes, of which 250 bytes are taken for header information. Record locations within the index use a 4-byte rowid. Assume that reading a page into memory takes 150 msec, and that the time needed for any query can be approximated by the time spent doing disk I/O.

3 levels
in total
计算方法

24 bytes

$$3846 / 24 = 160$$

$$100000 \div 160$$

$$= 625$$

- 1a) Calculate the space needed for the records/data of Rel1.

Hint: How much space is used by a single record? How many records per page? How many pages for the table? Then convert this back to space. [Note that each record is not split across pages, and each page has a header.]

$$625 \times 4096 = 2,560,000 \text{ bytes}$$

- 1b) Calculate the time taken to perform a table scan (i.e., linear scan) through the table Rel1.

Hint: How many pages are needed to make up the space occupied by the table? How many pages will be read from disk during a table scan?

$$625 \times 15\% = \text{seconds}$$

如果
occupancy
rate 只有

- 1c) Calculate the time taken, using the primary index, to execute the following query. Assume the selectivity of range condition on B is 10%.

SELECT C
FROM Rel1
WHERE A = 'AQG' AND (B BETWEEN 'WPQ' AND 'XYZ');

↓ 10% of whole B

X1 Count down
因为是 average 直接求在
这

Exercise 2. Index creation with PostgreSQL

因为 A 为 AQG 的时候，已经只有 1000 个 B 和匹配
至于 B 为 PK 时 record 与不满足 100-00

Find which indices already exist for the tables you defined in early tutorials and assignments.

Identify an extra index, which would be useful for one of the queries you wrote (supposing that the database became much bigger, with many more records in each table!). Create this index.

Note: To check which indexes are present in your schema (plus some details), you can use the following SQL commands in PostgreSQL:

`SELECT * FROM PG_INDEXES;`

↓ using RowC)

time:

$$150 \times 3 = 150 \times 4$$

$$450ms - 600ms$$

最少才 46

只有 162 页 3 页

1 : 160 : 1 ←

100 < 160

suffix index

in the same pose;

但是 Data 不是

最多

1 : 160 : 1 ←

最少 2 页

Q 609 if ~~for~~ intermediate level number: for B+ Tree

for index row, we only need

(A, B, rowId) pointer

$$4 + 12 + 4 = 20$$

mention in question

of pointer =
page number in
next layer

最后还要读 page

150 ms

$$\frac{3846}{20} = \lceil 192.3 \rceil = 192 \text{ Entry / page}$$

$$\frac{625}{192} = \lceil 3.2 \rceil = 4 \text{ pages in } \swarrow \text{ index layer}$$

$$\lceil \frac{4}{192} \rceil = 1 \text{ page in root layer}$$

在 page number 也是

(A, B, rowId) = 20 bytes

