Name:

Fasi-ur-Rehman

Reg No.:

2022-CS-71

Section:

B

Assignment:

Problem Set 2

Submitted to:

Sir Samyan

Problem Set #2

Problem 1

Question 1.1

$= 3 + 4 + 10 + 5 + 4 + 2 + 4 + 224$

Total row size = 256

Question 1.2

Disk Block size = 64 KB
= (64 × 1000) MBytes
= 64000 Bytes

No. of rows = Total space / Row size
= 64000 / 256

No. of rows = 250

Question 1.3

DB block size = 64 MB
= 64000000 Bytes

No. of rows = $\frac{64000000}{256}$

No. of rows = 250000

Question 1.4

Rows per quarter = Total students × Classes per student per quarter × Percentage writing evaluations

Rows per quarter = 15000 × 3 × 0.5 = 22500

Total rows after 40 quarter = 22500 × 40
= 900000.

∵ Size of single Rows = 256 bytes

Total size of table = Rows × size of row
= 900000 × 256
= 230400000 bytes

Total size in MB = $\frac{230400000}{1000000}$

Total size in MB = 230.4

Question 1.5

Size of DB block = 64 MB
Size of Table = 230.4 MB

No. of DB blocks = $\frac{230.4}{64}$

No. of DB block = 3.6 DB block.

- It will be rounded up to 4 DB block as DB block can not be in Fractions.

Question 1.6

Seek Time = 10 ms
Size of row = 256 bytes
Transfer speed = 100 MB/sec

$$\text{Transfer time} = \frac{\text{Size of row}}{\text{Transfer speed}}$$

$$\text{Transfer time} = \frac{256 \times 10^{-6}\ \text{MB}}{100\ \text{MB/sec}}$$

~~Transfer time = 0.00256 seconds~~ ms.

$$\text{Transfer time} = 2.56 \times 10^{-6}\ \text{seconds}$$

$$\text{Transfer time} = (2.56 \times 10^{-6} \times 1000)\ \text{ms}$$

$$\boxed{\text{Transfer time} = 0.00256\ \text{ms}}$$

$$\text{Total time} = 10\ \text{ms} + 0.00256\ \text{ms}$$
$$= 10.00256\ \text{ms}$$

$$\text{Total time in Hour} = \frac{10.00256}{3600000} = 0.00000278$$
$$= 0.00000278\ \text{hours}$$

- Round off value = 0.0 hours

Question 1.7

Transfer time for a disk block = $\frac{\text{Disk block size}}{\text{Transfer speed}}$

$= 64\text{KB} / 100\text{MB/sec}$

Transfer time per disk block = 0.00064 seconds

Total time to read single row = Seek time + $\frac{\text{Transfer time per disk block}}{\text{Number of rows.}}$

$= 10\text{ms} + \frac{0.00064}{250}$

$= 10.000000256\text{ ms}$

Total time = 10.000000256 = 0.010s

Problem 2

Question 2.1

Maximum time For Finding a record in Full scan = $\frac{0.5 \times \text{Total Data}}{\text{transfer speed}}$

$= \frac{0.5 \times 3200\text{ GB}}{10\text{ MB/sec}}$

Max time = 16000 seconds

Average Response time = seek time + $\left(\frac{\text{Max time}}{\text{No. of rows}}\right)$

Average response time $= 10ms + \left(\frac{16000}{5 \times 10^7}\right)$

$= 10ms + 0.00032.s = 0.010 seconds$

Average response time = 0.010 seconds

Question 2.2:

Given that 1% of table rows are responsible for 90 % of query traffic. It means these rows are most likely to be queried. So, we can improve by pulling it in RAM. As RAM is much faster. So, we can we can improve the response time by implementing Cache. i.e by putting 1% row which are mostly accessed in cache So, it can be fetched quickly.

Estimated Improvement:

while the exact improvement depend on specific cache strategy and access patterns, By implementing this average time will be greatly reduced. as compared to average without cache.

Problem 3

Question 3.1

Lets assume data types

user id : int32 (4 Byte).
user name : char[64] (64 Bytes)
item id : int32 (4 Byte)
item name : char [64] (64 Byte)
transaction id : long [8 Byte)
amount of money : double (8 byte)

Size of row = 4+64+4+64+8+8 = ~~156 Byte~~
Size of row = 152 Byte.

Question 3.2

user id : int
user name : char
item id : int
item name : char
transaction id : long
amount of money : double

- Most apporopriate data type For User ID is ~~char[64]~~ which is actually string. int [32]

Question 3.3

Most appropriate data type For ~~item~~ user name is char [64]

Question 3.4

Most appropriate data type for item id is int [32]

Question 3.5

Most appropriate data type for Item Name is char [64]

Question 3.6

Most appropriate data type for transaction ID is Long.

Question 3.7

Most appropriate data type for Amount of Money is Double as it can be in decimal

Question 3.8

Total Size of Table = 152 × 1 trillion

Size of table in TB = $\frac{152 \times 10^{12}}{2^{40}}$

= 138.24 TB

Problem 4

Question 4.1

RAM transfer speed = 100 GiB/sec

Size of Table = 200 TiB = 200×10^{40} Bytes

Time to read table = $\frac{\text{Size of table}}{\text{Speed of RAM}}$

$= \frac{200 \times 10^{40}}{100 \times 10^{30}}$

$= \frac{200 \times 10^{10}}{100}$

$= 2 \times 10^{10}$

Time to read table = 2048 seconds

Question 4.2

Disk transfer speed = 100 MB/sec

Disk blocks size = 64 KiB

Size of Table = 200×2^{40} Bytes

∴ Average response time as calculated in 2.1 is 0.010 seconds.

Time per Row = 0.010 seconds.

Time to Read Table From Disk = Time Per Row × Number of Rows.

Time to read Table from Disk = 0.010×10^{12}

Time in Days = $\frac{0.010 \times 10^{12}}{60 \times 60 \times 24}$

= 11.57 days

Question 4.3

Number of DB blocks = $\frac{\text{Size of table}}{\text{Size of DB block}}$

No. of DB blocks = $\frac{200\,000000 \text{ MB}}{64 \text{ MB}} = 3125000$

Time to read 1 DB block = $\frac{\text{Size of DB block}}{\text{Disk transfer speed}}$

Time to read 1 DB block = $\frac{64 \text{ MB}}{100 \text{ MB/sec}} = 0.64 \text{ sec}$

Total time to read table = No. of DB block x Time to read 1 block

Total time to read table = 3125000 x 0.64

Total time to read table = 2000000 sec

Converting in In Days :

= $\frac{2000000}{24 \times 60 \times 60}$ = 23.14 Days

Rounded off to 23 Days

Question 4.4

Total cost = Total RAM required x Cost per TB

Total RAM Required = 200 TB

Cost per TB = $6000

Total cost = 200 × 6000

= $1200000

Question 4.5

Total Cost = Total disk space × Cost per TB

= 200 TB × $100 $/TB

Total Cost = 20000 $

Problem 5

Question 5.1

For e-comerse site I think following tables are esential.

User Data: This table will have all the information about users registered. It might also have their credit card details if saved by user.

Products: This table will have the data about all the products that are listed on site For sale. Their status like in stock or not. and many more such attributes.

Question 5.2

Number of bits to store Product Id can be calculated by:

Number of bits = $\lceil \log_2 (\text{No. of Product ID}) \rceil$

Number of bits = $\lceil \log_2 (2 \times 10^9) \rceil$

Number of bits = 32.89 ≈ 33

Data types for provided data will be **big int** (int64, — 8 bytes)
as int32 < 33 < int64

Question 5.3

For storing unique ID of 1 Billion users.

No. of Bits = $\lceil \log_2 (10^9) \rceil$

No. of Bits = 29.89 ≈ 30

As we need to store 30 bits we can use **int** (int32 — 4 bytes)

Question 5.4

Order ID : int 64
Product ID : big int (int 64 - 8 bytes)
User ID : int 32
Quantity : int 32
Timestamp : 4 Bytes
IP address : 4 Bytes
Mailing address: char[100]

size of row = 8 + 8 + 4 + 4 + 4 + 100 = **132 bytes**

Question 5.5

Total size = Size of 1 row X Number of orders per day X Number of days

Total size = 132 bytes X (100×10^6) X 7

Total size = 9.264×10^{10} Bytes

Total size = 92.64×10^9 Bytes

Total size = 92.64 GB

Question 5.6

RAM seek time = 20 ns

Size of table = 10 GB = 10×10^9 Bytes

Time = RAM seek time X no. of seeks

no. of seeks = $\frac{\text{Total size}}{\text{Size of record}} = \frac{10 \times 10^9}{132}$

No. of seeks = 75757575.76

time = 20 ns X 75757575.76

Time in ms = $\frac{1515151515}{1000000}$

= **1515.15152 ms**

Question 5.7

Disk access time = 10 ms

size of table = 10 GB = 10×10^9 Bytes

Time = Disk time X no. of seeks

No. of seeks = $\frac{10 \times 10^9}{132}$

= 75757575.76 seeks ~~seconds~~

Time = 10 ms X 75757575.76

= ~~(10 000000 ns) X 75757575.76~~

In Days

Time = $\frac{10 \text{ ms} \times 75757575.76}{86400000 \text{ ms/day}}$

Time = 8.76 Days

Question 5.8

Disk access = 10ms

Size of each block = 64 mb

Total size of table = 10GB = 10000MB

No. of blocks = $\frac{10000}{64}$ = 156.25 blocks

Records per block = $\frac{64000000 \text{ Bytes}}{132}$

Records per block = 484848.4848

Number of seeks = $\frac{\text{No. of records}}{\text{Records per block}}$

$= \frac{10 \times 10^9 \text{ Bytes}}{484848.4848}$

Number of seeks = 20625

Time = Disk access X No. of seeks

Time = 10 ms X 20625

Time = 206250 sec = 206 ms

Question 5.9

If we parallelized perfectly n machines then lookup time for a record will be reduced. It can be expressed as

$$T_n = \frac{T_1}{n}$$

- where T_n is time for looking up when n machines are parallelized.
- T_1 is time for looking up record when looked up using 1 machine.
- For 10 machines

$$T_{10} = \frac{T_1}{10}, \quad \text{Speedup} = \frac{T_1}{T_n}$$

$$\text{speed up} = \frac{T_1}{\frac{T_1}{10}} = 10$$

So, 10 machines will look up 10 times faster.

Question 5.10

Size of Data = 10GB = 10×10^9 Bytes

Network transfer time = $\frac{10 \times 10^9 \text{ Bytes}}{100 \times 10^7 \text{ Bytes/sec}}$

$= \frac{1}{10}$ sec

- In Milliseconds

$$\text{Network transfer time} = \frac{1}{10} \times 1000 = 100 \text{ ms}$$