Duy Truong
Homework 3.

94+3

12.1

We have tuples (kangaroo, 17) though (baboon, 12) using tuple number t_1 to t_{12}. We refer to the j^{th} run used by the i^{th} pass r_{ij}. We sorted runs have three blocks each they ar

$r_{11} = \{t_3, t_1, t_2\}$
$r_{12} = \{t_6, t_5, t_4\}$
$r_{13} = \{t_9, t_7, t_8\}$
$r_{14} = \{t_{12}, t_{11}, t_{10}\}$

> merge = r_{21}

> merge = r_{22}

Each pass merges 3 runs. So we have

$r_{21} = \{t_3, t_1, t_6, t_9, t_5, t_2, t_7, t_4, t_8\}$? — 1

$r_{22} = ?$

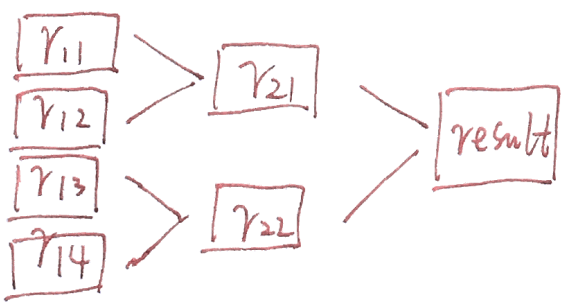
Last one is second pass, the tuples are completely sorted into one run.

$r_{31} = \{t_{12}, t_3, t_{11}, t_{10}, t_1, t_6, t_9, t_5, t_2, t_7, t_4, t_8\}$

$r_{21} = \{t_3, t_1, t_6, t_5, t_2, t_4\}$.

$r_{22} = \{t_{12}, t_{11}, t_{10}, t_9, t_7, t_8\}$.

Refer to the book for sort-merge.



- salary > 60.
- B+ tree

The size of a tuple is $20 * 5 = 100$

A page size is 4096 bytes

The number of tuples per page is $\frac{40.96}{100} = 40$

The number of pointer per page is $\frac{4096 - 4}{20 + 4} + 1 = 171$

The high of B^+ tree is $\lceil \log_{\frac{171}{2}} 25000 \rceil = 2$.

The number of leaf pages is $1000000 / (\frac{171 - 1}{2}) * \frac{100 - 60}{100 - 51}$

$= 9604.$

The number of qualifing tuples $= 1000000 \times \frac{100 - 60}{100 - 51}$

$= 816327$

$Cost = (2-1) + 9604 + 816327 = 825933$

- File scan.

Size of tuple $= 100$.

The number of fit in a page $\frac{4096}{100} = 40$

$Cost = 25000$

$\Rightarrow$ Cost $= 25000$

age = 25.

The number of qualifitid tuple $1000000 \times \frac{1}{50} = 20000$

Cost = $1.2 \times 20000 = 24000$

File scan.

qualifying leaf page is $1000000 \times \frac{1}{70-21} * 140 = 510$.

The number of pointer is $\frac{4096-4}{20 \times 2 + 4} + 1 = 94$

The height of tree is $\lceil \log_{\frac{94}{2}} 250000 \rceil = 2$.

Cost = $510 + 2 = 512$.

$\Rightarrow$ best cost = 512 (B+ tree (age, salary))

age > 25.

numbe of qualifying leaf is $1000000 * \frac{(70-25)}{70-21} / 40 = 22960$.

number of pointer $= \frac{4096-4}{20 \times 2+4} + 1 = 94$

the height of tree $= \log_{\frac{94}{2}} 25000 = 2$.

Cost $= 2 + 22960 = 22962$.

4) eid = 1234567890.

Hash index eid.

The number qualifying tuples is 1.

$\Rightarrow$ Cost $= 1.2 * 1 = 1,2$

5) Salary > 60 and age = 25.

select * from Employees where salary > 60 and age = 25

cluster B+ tree (age, salary)

cost = 512 (question 2)

unclustered B+ tree on salary

cost = 825923 (question 1)

$\Rightarrow$ the best cost = 512

6> Salary > 60k and age > 25

Salary > 60k ~~cost~~ cost = 825933 (unclustered B+tree)

age > 25 cost = 22962 (clustered B+tree on salary)

=> best cost = 22962.

7> Salary > 60k and title = manager

B+tree on salary

cost = 825933

8> Salary > 60k, age > 25 and title = manager

clustered B+ tree on (age, salary)

cost = 512 Result: same as "6>" -2

This question is: Salary > 60k,

age > 25 and title = manager.

△ (not "=")

So it should be the same as question No. 6

That's why you lose point.