

Yao_Xiao_hw4_report

Q1

Name
Austria
Belgium
Switzerland
Denmark
Finland
Greece
Netherlands
Norway
Poland
Portugal
Russian Federation
Sweden

Country	Capital
Anguilla	The Valley
Antigua and Barbuda	Saint John's
Aruba	Oranjestad
Bahamas	Nassau
Barbados	Bridgetown
Belize	Belmopan
Bermuda	Hamilton
Canada	Ottawa
Cayman Islands	George Town
Costa Rica	San José

only showing top 10 rows

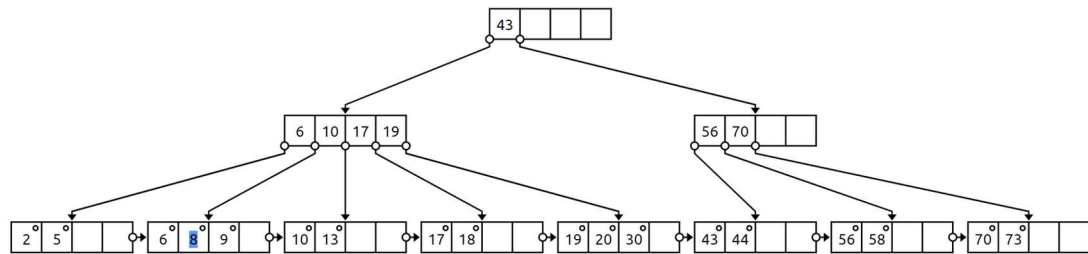
Country	official_languages
Anguilla	English
Antigua and Barbuda	English
Aruba	Dutch
Barbados	English
Belize	English
Bermuda	English
Canada	English, French
Cayman Islands	English
Costa Rica	Spanish
Cuba	Spanish

Continent	avg_life_expectancy
Europe	75.01
Africa	61.75555555555555
North America	73.85555555555555
South America	71.675
Oceania	78.8
Asia	70.10689655172413

Continent	cnt
Europe	5
Africa	5
North America	5
Oceania	3

Q2

A



– Finding 10 in the range

43 → 6, 10, 17, 19 → 10, 13

This step costs 3 I/O's

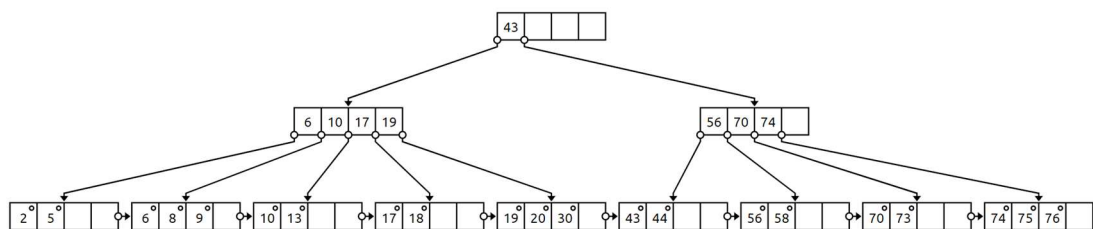
– Then sequential traversal of leaves until find a block is larger than 60

17, 18 → 19, 20, 30 → 43, 44 → 56, 58 → 70, 73

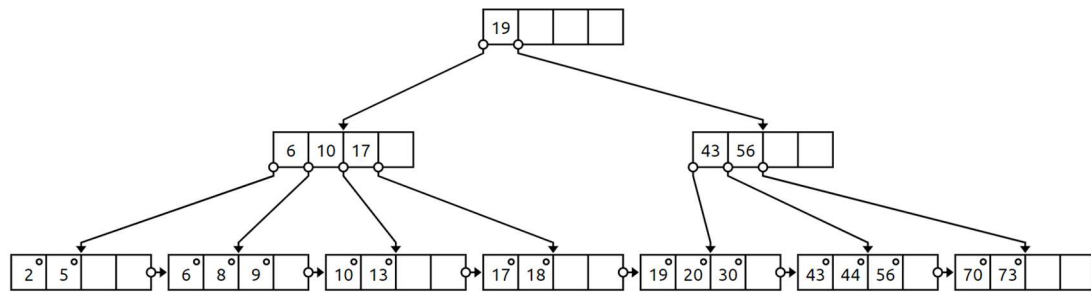
This step costs 5 I/O's

So totally we need 8 I/O's

B



C



Q3

A

$$\text{cost}(\text{outer} = R) = B(R) + \frac{B(R)}{M-2} * B(S) = 500 * \frac{500}{102-2} * 1000 = 5500$$

Cost:

- Read R once: cost $B(R)$
- Outer loop runs $B(R)/(M-2)$ times, and each time need to read S: costs $B(R)B(S)/(M-2)$
- Total cost: $B(R) + B(R)B(S)/(M-2)$

for each $(M-2)$ blocks b_r of R

do for each block b_s of S

do for each tuple r in b_r

do for each tuple s in b_s

do if r and s join then output (r, s)

B

$$B(R) + B(S) = 1500 \leq M^2$$

Step 1: split R into runs of size 100 which takes 5 runs, then split S into runs of size 100 which takes 10 runs.

Cost: $2B(R) + 2B(S)$

Step 2: merge 5 runs from R and 10 runs from S; output a tuple on a case by cases basis

Total cost: $3B(R) + 3B(S) = 4500$

C

Step 1

Hash S into 100 buckets and then send all buckets to disk

Step 2

Hash R into 100 buckets and then send all buckets to disk

Step 3

Join every pair of corresponding buckets

Total cost: $3B(R) + 3B(S) = 4500$

D

Loading R and then iterate over R, for each tuple, fetch corresponding tuple(s)

from S. Join R and S.

$$\text{cost} = B(R) + \frac{B(S)}{V(S,a)} * T(R) = 500 + \frac{1000}{20} * 10000 = 500500$$