Radhika Patwari
18CS10062

22/3/2022

Class Test 2

Question 1

list = 110, 401, 1169, 16273

gaps: 291, 768, 15104

VB code

110 - 0
55 - 1
27 - 1
13 - 1
6 - 0
3 - 1
1

11101110

291 → 00000010 10100011

768 → 00000110 10000000

15104 → 01110110 10000000

list	110	401	1169	16273
gap		291	768	15104
VB code	11101110	00000010 10100011	00000110 10000000	01110110 10000000
Gamma code	00011111 10101110	00000001 11111110 00100011	00011111 11101000 00000000 00000111 11111101 00000000	00000111 11111111 11011011 00000000

18CS10062
Radhika Patwari

Question 4

(a) KT

(b)

18CS10062
Radhika Patwari

Question 3

(a) Tokens = 30,000,000

total bytes for all tokens = 8 × 30,000,000

terms = 1,50,000

total byte for all terms = 4 × 1,50,000

∴ Total bytes = 240,000,000
+ 4,200,000
244,200,000

Given, transfer time per byte = 2×10^{-8} s

For 15 blocks, avg seek time = 5×10^{-3} s

∴ Total time = $5 \times 10^{-3} + (2 \times 10^{-8})(244,200,000)$

$= 5 \times 10^{-3} + 4.88$

$= 0.005 + 4.88$

$= 4.885$ sec

(b) Compare & swap word speed = 10^{-8} sec

∴ Total sorting time = $10^{-8} \times (1,50,000) \log(1,50,000)$

(c) $5 \times 10^{-3} + (2 \times 10^{-8})(244,200,000)$

$= 4.885$ sec

18CS10062
Radhika Patwari

(d) Map:

Collection → (<term1, list(docID)>,
<term2, list(docID)>,
<term3, list(docID)> ...)

Reduce: (<term1, (doc1, doc2)> → (<term1, doc1:1
<term1, (doc3, doc2)> doc2:2
⋮ doc3:1>,
) ⋮
)

(e) MapReduce Algorithm can handle these cases.
We can divide the posting list of terms into 100 GB data using an initial preprocessing map function.