CZ4031 Database System Principles

Project 1: B+ Tree

Documentation

Group 37:

NG YUE JIE ALPHAEUS (U2021469L)

TAN PHENG KHAI (U2020088C)

KEVIN LIU KAI (U1921021F)

SANKAR SAMIKSHA (U2021021D)

TRAN TRUNG DUNG (U2020441G)

1. Contribution of Each Member **3**

2. Design of the storage component **3**

2.1 How each data item is stored as a field 3

2.2 How fields are packed into a record 4

2.3 How records are packed into a block 4

3. Design of the B+ tree component, including the data structure of a node and the maximum number of keys a node maintains. **4**

3. Results of the experiments in Part 2 **5**

Block Size: 200B 5

Block Size: 500B 5

4. Installation guide **6**

5. Appendix: **7**

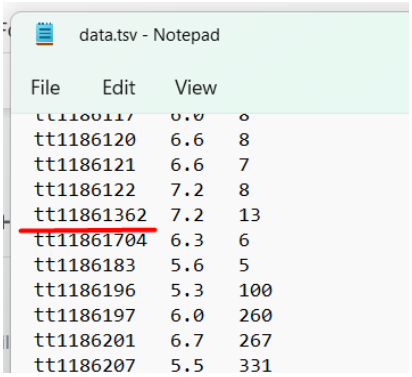
1. Contribution of Each Member

Name	Contribution
NG YUE JIE ALPHAEUS	Made the Documentation. Implemented the Code.
TAN PHENG KHAI	Made the Documentation. Implemented the Code.
KEVIN LIU KAI	Made the Documentation. Implemented the Code.
SANKAR SAMIKSHA	Made the Documentation. Implemented the Code.
TRAN TRUNG DUNG	Made the Documentation. Implemented the Code.

2. Design of the storage component

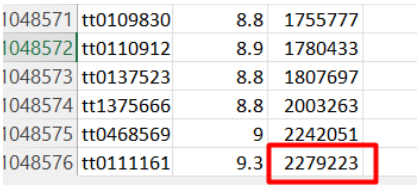
2.1 How each data item is stored as a field

11B reserved for tconst in char data type.



Longest is ttxxxxxxx, 10 Characters. Total bytes to be used to store tconst will be 10 (data) + 1 (/0) = 11B of character data type.

4B reserved for numVotes in unsigned int data type.
4B reserved for averageRating float data type.



1048571	tt0109830	8.8	1755777
1048572	tt0110912	8.9	1780433
1048573	tt0137523	8.8	1807697
1048574	tt1375666	8.8	2003263
1048575	tt0468569	9	2242051
1048576	tt0111161	9.3	2279223

numVotes
The data was sorted and the largest value found was 7 digits in length, unsigned short int which is 4B, 2^32 of a max value of 4294967296 is sufficient to handle the numVote record.

averageRating
The average rating min value is 0, max is 10. Float of 4B enough to capture the data.

Fixed Format with Fixed Length

tconst (11B, Char)	averageRating (4B, float)	numVotes (4B, unsigned int)
tt0000001	5.6	1645
tt0000002	6.1	198
. . tt11861362	. . 7.2	. . 13

Relational data

Figure 1: A figure to illustrate the design and size information of field and record

2.2 How fields are packed into a record

Fixed length was chosen. One record contains 11B of char data + 4B of float data + 4B of unsigned int data, totalling 19B. We decided to use fixed length as the storage efficiency from implementing variable length would be negligible or actually worse than a fixed length.

To note: However, in the actual implementation, the c++ compiler would add a 1B padding to the record struct, making it 20B. However, when calculating the final amount of size and blocks used, the original 19B space is used as confirmed by Prof Long Cheng.

Rationale for choosing fixed length

Since averageRating and numVotes are not char data types, a fixed float and unsigned int data type is sufficient for them. The field that would potentially benefit from the variable length implementation would be tconst.

However, to consider implementing variable length for tconst, it would only be of benefit if the tconst has a range of value. However, the tconst data type is minimally 9B or longer, which means a header of 4B unsigned int would result in minimally 13B, which is already more than the original 11B allocated via fixed length.

Additional thoughts

The experiment stated that the tconst value should be saved as a string, if however we could assume that all data would have the same “tt” in the first two byte, and we save the remaining “0000001” as an unsigned int instead, we would be able to save space by reducing it from an 11B char data to a 4B unsigned int data. Increasing storage efficiency.

In total, 19B is needed to store one record with Fixed-length.

2.3 How records are packed into a block

Records are fitted 19B at a time into a block, in a contiguous fashion. At the 190th byte in a 200B block, we're able to store the first 10B in the first block, and the remaining 9B in the 2nd block. Thus, increasing space efficiency by utilising all the data in a block.

We are able to do this because after importing in the data, we sort it by each record's numVote. Afterwards, our query/insert/delete algorithm works on the premise that data is stored in a physically contiguous fashion, and accesses each record as such. If a record is split into 2 separate blocks, 2 I/Os are taken into account.

3. Design of the B+ tree component, including the data structure of a node and the maximum number of keys a node maintains.

Data Structure of a node:
A node of the B+ tree contains

1) Metadata (127 bits)
    a) Bool isLeaf (1 bit)
    b) Int currentKeySize (32 bits)
    c) Int currentPointerSize (32 bits)
    d) Int maxKeySize (32 bits)
    e) Int maxPointerSize (32 bits)
    f) Struct Address addressInDisk (80 bits)
2) Array of keys (n x 32 bits)
3) Array of pointers to children nodes ((n+1) x 64 bits)

Depending on the size of the block allocated, 200B or 500B, the node's maximum number of keys(n) is dynamically calculated.

$$Size\ of\ Node\ =\ size\ of\ metadata\ +\ size\ of\ key\ array\ +\ size\ of\ array\ of\ pointers$$
$$=\ 127\ +\ 80\ +\ 32n\ +\ (n\ +\ 1)\ \times\ 64$$
$$=\ (271\ +\ 96n)\ bits$$

$$Therefore,\ for\ 200B\ blocks,\ each\ block\ can\ store:$$
$$Size\ of\ Node\ =\ 200\ \times\ 8\ bits\ =\ 1600\ bits$$
$$n\ =\ \frac{1600-271}{96} = 13.84\ =\ 13\ entries\ (Rounded\ down)$$

$$For\ 500B\ Blocks,\ each\ block\ can\ store:$$
$$Size\ of\ Node:\ 500\ \times\ 8\ bits\ =\ 4000\ bits$$
$$n\ =\ \frac{4000-271}{96} = 38.84\ =\ 38\ entries\ (rounded\ down)$$

3. Results of the experiments in Part 2

Block Size: 200B
Experiment 1 Results:
----------Experiment 1---------- Number of Records : 1070319 Total number of blocks : 112389 Actual size of database : 21514680
Experiment 2 Results:
----------Experiment 2---------- ******BPTREE DETAILS****** Parameter n of B+ Tree (number of keys in Node): 13 Total NodeSize: 5530 nodes height of B+ Tree is: 5 height of B+ Tree is: 5 Content Of Root Node: ********** Node's Disk address: 0x7fe0e302a220 currentKeySize :4 currentPointerSize :5 node type :InternalNode max Key Size: 13 keyArray: [3591,7911,16859,47952,0,0,0,0,0,0,0,0,0,] PointerArray: [0x7fe0e302a090,0x7fe0e3063368,0x7fe0e309c578,0x7fe0e30d5788,0x7fe0e30d5850,0,0,0,0,0,0,0,0,0,] ********** Content of First Child Of Root Node ********** Node's Disk address: 0x7fe0e302a090 currentKeySize :7 currentPointerSize :8 node type :InternalNode max Key Size: 13 keyArray: [453,901,1349,1797,2245,2693,3141,0,0,0,0,0,0,] PointerArray: [0x7fe0e2fcd368,0x7fe0e2fd4640,0x7fe0e2fdb850,0x7fe0e2fe2a60,0x7fe0e2fe9c70,0x7fe0e2ff0e80,0x7fe0e2ff8090,0x7fe0e2fff2a0,0,0,0,0,0,0,0,] **********
Experiment 3 Results:

```
----------Experiment 3----------
Resetting block access before carrying on with the experiment
Retrieving movies with 'numVotes' equal to 500:

Searching~~~~~~~~~~~~~~~~~~~~~~~~~~~~
**********Index Node Content**********
index root Node accessed:
**********
Node's Disk address: 0x7fe0e302a220
currentKeySize :4
currentPointerSize :5
node type :InternalNode
max Key Size: 13
keyArray: [3591,7911,16859,47952,0,0,0,0,0,0,0,0,0,]
PointerArray:
[0x7fe0e302a090,0x7fe0e3063368,0x7fe0e309c578,0x7fe0e30d5788,0x7fe0e30d5850,0,0,0,0,0,0
,0,0,0,]
**********
2nd index block accessed:
**********
Node's Disk address: 0x7fe0e302a090
currentKeySize :7
currentPointerSize :8
node type :InternalNode
max Key Size: 13
keyArray: [453,901,1349,1797,2245,2693,3141,0,0,0,0,0,0,0,]
PointerArray:
[0x7fe0e2fcd368,0x7fe0e2fd4640,0x7fe0e2fdb850,0x7fe0e2fe2a60,0x7fe0e2fe9c70,0x7fe0e2ff0e
80,0x7fe0e2ff8090,0x7fe0e2fff2a0,0,0,0,0,0,0,0,]
**********
3nd index block accessed:
**********
Node's Disk address: 0x7fe0e2fd4640
currentKeySize :7
currentPointerSize :8
node type :InternalNode
max Key Size: 13
keyArray: [509,565,621,677,733,789,845,0,0,0,0,0,0,0,]
PointerArray:
[0x7fe0e2fc8b88,0x7fe0e2fc9998,0x7fe0e2fca7a8,0x7fe0e2fcb5b8,0x7fe0e2fcc3c8,0x7fe0e2fcd1
d8,0x7fe0e2fce240,0x7fe0e2fcf050,0,0,0,0,0,0,0,]
**********
4nd index block accessed:
**********
Node's Disk address: 0x7fe0e2fc8b88
currentKeySize :7
currentPointerSize :8
node type :InternalNode
max Key Size: 13
keyArray: [460,467,474,481,488,495,502,0,0,0,0,0,0,0,]
PointerArray:
[0x7fe0e2fc7418,0x7fe0e2fc75a8,0x7fe0e2fc7738,0x7fe0e2fc78c8,0x7fe0e2fc7a58,0x7fe0e2fc7b
e8,0x7fe0e2fc7f08,0x7fe0e2fc8098,0,0,0,0,0,0,0,]
**********
5nd index block accessed:
```

****
Node's Disk address: 0x7fe0e2fc7f08
currentKeySize :7
currentPointerSize :8
node type :LEAF
max Key Size: 13
keyArray: [495,496,497,498,499,500,501,0,0,0,0,0,0,]
PointerArray:
[0x7fe0e2e01494,0x7fe0e2e01d04,0x7fe0e2e02574,0x7fe0e2e02eac,0x7fe0e2e037f8,0x7fe0e2e
04054,0x7fe0e2e048ec,0,0,0,0,0,0,0x7fe0e2fc8098,]
****

All records found within first 5 data blocks: 0x7fe0e2e04054
**********Record Content 1 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 5.1
tconst for current record: tt0082841
**********Record Content 2 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 4.1
tconst for current record: tt1727519
**********Record Content 3 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 5.5
tconst for current record: tt6038946
**********Record Content 4 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 8
tconst for current record: tt4392820
**********Record Content 5 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 7.1
tconst for current record: tt0082275
**********Record Content 6 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 8
tconst for current record: tt4627488
**********Record Content 7 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 7.4
tconst for current record: tt1632756
**********Record Content 8 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 7.8

```
tconst for current record: tt1857596
**********Record Content 9 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 7.7
tconst for current record: tt1640740
**********Record Content 10 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 5.3
tconst for current record: tt1684558
**********Record Content 11 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 7.1
tconst for current record: tt0720140
**********Record Content 12 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 4.6
tconst for current record: tt0119415
**********Record Content 13 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 7.8
tconst for current record: tt3916858
**********Record Content 14 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 7.1
tconst for current record: tt3440780
**********Record Content 15 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 6.3
tconst for current record: tt0218000
**********Record Content 16 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 7.3
tconst for current record: tt3263598
**********Record Content 17 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 9.2
tconst for current record: tt8044824
**********Record Content 18 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 8.7
tconst for current record: tt0588114
**********Record Content 19 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 7.1
```

tconst for current record: tt6985094
**********Record Content 20 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 6.4
tconst for current record: tt11691696
**********Record Content 21 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 9.1
tconst for current record: tt0514442
**********Record Content 22 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 8.2
tconst for current record: tt4713652
**********Record Content 23 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 4.6
tconst for current record: tt6318954
**********Record Content 24 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 6
tconst for current record: tt1340802
**********Record Content 25 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 7.2
tconst for current record: tt0647512
**********Record Content 26 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 7.2
tconst for current record: tt7431738
**********Record Content 27 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 5.1
tconst for current record: tt0289115
**********Record Content 28 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 7.7
tconst for current record: tt9614612
**********Record Content 29 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 8.1
tconst for current record: tt7517336
**********Record Content 30 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 7.6

tconst for current record: tt0716912
**********Record Content 31 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 8.1
tconst for current record: tt0163456
**********Record Content 32 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 8.2
tconst for current record: tt0640308
**********Record Content 33 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 6.7
tconst for current record: tt0052815
**********Record Content 34 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 4.7
tconst for current record: tt0062345
**********Record Content 35 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 7.9
tconst for current record: tt4247300
**********Record Content 36 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 6.4
tconst for current record: tt4038966
**********Record Content 37 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 8
tconst for current record: tt0640298
**********Record Content 38 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 8.6
tconst for current record: tt0588214
**********Record Content 39 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 8.1
tconst for current record: tt0517623
**********Record Content 40 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 6.4
tconst for current record: tt0327546
**********Record Content 41 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 8.4

tconst for current record: tt7860116
**********Record Content 42 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 7.9
tconst for current record: tt1371697
**********Record Content 43 at address: 0x7fe0e2e04054 **********
Record address in disk: 0x7fe0e2e04054
NumVotes for current record: 500
Average Rating for current record: 3.7
tconst for current record: tt0054298

Summarised Information:
total number of records accessed: 43
Average avgRating of the records: 7.01163

Number of index blocks accesses : 5
Number of record blocks accesses : 5

Experiment 4 Results:

----------Experiment 4----------
Resetting block access before carrying on with the experiment
Retrieving those movies with the attributes 'numVotes' from 30K to 40K

Searching~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
**********Index Node Content**********
index root Node accessed:
****
Node's Disk address: 0x7fe0e302a220
currentKeySize :4
currentPointerSize :5
node type :InternalNode
max Key Size: 13
keyArray: [3591,7911,16859,47952,0,0,0,0,0,0,0,0,0,]
PointerArray:
[0x7fe0e302a090,0x7fe0e3063368,0x7fe0e309c578,0x7fe0e30d5788,0x7fe0e30d5850,0,0,0,0,0,0,0,0,0,]
****
2nd index block accessed:
****
Node's Disk address: 0x7fe0e30d5788
currentKeySize :7
currentPointerSize :8
node type :InternalNode
max Key Size: 13
keyArray: [18722,20935,23519,26868,30797,35447,41169,0,0,0,0,0,0,]
PointerArray:
[0x7fe0e3078998,0x7fe0e307fba8,0x7fe0e3086db8,0x7fe0e308dfc8,0x7fe0e30951d8,0x7fe0e309c3e8,0x7fe0e30a3788,0x7fe0e30aa998,0,0,0,0,0,0,]
****
3nd index block accessed:

****
Node's Disk address: 0x7fe0e30951d8
currentKeySize :7
currentPointerSize :8
node type :InternalNode
max Key Size: 13
keyArray: [27248,27749,28276,28796,29315,29869,30395,0,0,0,0,0,0,]
PointerArray:
[0x7fe0e30897e8,0x7fe0e308a5f8,0x7fe0e308b408,0x7fe0e308c218,0x7fe0e308d028,0x7fe0e30
8de38,0x7fe0e308edd8,0x7fe0e308fbe8,0,0,0,0,0,0,]
****
4nd index block accessed:
****
Node's Disk address: 0x7fe0e308edd8
currentKeySize :7
currentPointerSize :8
node type :InternalNode
max Key Size: 13
keyArray: [29949,29978,30041,30090,30177,30248,30333,0,0,0,0,0,0,]
PointerArray:
[0x7fe0e308d4d8,0x7fe0e308d668,0x7fe0e308d7f8,0x7fe0e308d988,0x7fe0e308db18,0x7fe0e30
8dca8,0x7fe0e308e158,0x7fe0e308e2e8,0,0,0,0,0,0,]
****
5nd index block accessed:
****
Node's Disk address: 0x7fe0e308d7f8
currentKeySize :7
currentPointerSize :8
node type :LEAF
max Key Size: 13
keyArray: [29978,29982,29988,29996,30022,30034,30037,0,0,0,0,0,0,]
PointerArray:
[0x7fe0e2fa6510,0x7fe0e2fa6524,0x7fe0e2fa6538,0x7fe0e2fa654c,0x7fe0e2fa6560,0x7fe0e2fa65
74,0x7fe0e2fa6588,0,0,0,0,0,0,0x7fe0e308d988,]
****

All records found within first 5 data blocks: 0x7fe0e2fa6560
**********Record Content 1 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30022
Average Rating for current record: 7.7
tconst for current record: tt0054167
**********Record Content 2 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30034
Average Rating for current record: 7.9
tconst for current record: tt0026778
**********Record Content 3 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30037
Average Rating for current record: 5.6
tconst for current record: tt0091828

**********Record Content 4 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30041
Average Rating for current record: 6.8
tconst for current record: tt3361792
**********Record Content 5 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30049
Average Rating for current record: 6.2
tconst for current record: tt1456941
**********Record Content 6 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30053
Average Rating for current record: 6.3
tconst for current record: tt0110443
**********Record Content 7 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30056
Average Rating for current record: 5.1
tconst for current record: tt0105629
**********Record Content 8 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30078
Average Rating for current record: 5
tconst for current record: tt0257516
**********Record Content 9 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30078
Average Rating for current record: 6.7
tconst for current record: tt0489664
**********Record Content 10 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30081
Average Rating for current record: 8.7
tconst for current record: tt2303687
**********Record Content 11 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30085
Average Rating for current record: 7.5
tconst for current record: tt2027128
**********Record Content 12 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30090
Average Rating for current record: 7.4
tconst for current record: tt0078718
**********Record Content 13 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30136
Average Rating for current record: 6.5
tconst for current record: tt0857265
**********Record Content 14 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30144
Average Rating for current record: 6.7
tconst for current record: tt1435513

**********Record Content 15 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30149
Average Rating for current record: 8.1
tconst for current record: tt2712740
**********Record Content 16 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30158
Average Rating for current record: 6
tconst for current record: tt0268397
**********Record Content 17 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30158
Average Rating for current record: 7.5
tconst for current record: tt0108211
**********Record Content 18 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30168
Average Rating for current record: 6
tconst for current record: tt9053874
**********Record Content 19 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30175
Average Rating for current record: 7.5
tconst for current record: tt8758202
**********Record Content 20 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30177
Average Rating for current record: 6.6
tconst for current record: tt0082533
**********Record Content 21 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30195
Average Rating for current record: 7.2
tconst for current record: tt1741256
**********Record Content 22 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30206
Average Rating for current record: 7.5
tconst for current record: tt3139086
**********Record Content 23 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30221
Average Rating for current record: 7.4
tconst for current record: tt0082089
**********Record Content 24 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30240
Average Rating for current record: 4.9
tconst for current record: tt0113253
**********Record Content 25 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30246
Average Rating for current record: 8.1
tconst for current record: tt0056111

**********Record Content 26 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30247
Average Rating for current record: 7.4
tconst for current record: tt4396630
**********Record Content 27 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30248
Average Rating for current record: 6.8
tconst for current record: tt7668518
**********Record Content 28 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30254
Average Rating for current record: 7.8
tconst for current record: tt0037382
**********Record Content 29 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30259
Average Rating for current record: 7.4
tconst for current record: tt0204993
**********Record Content 30 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30262
Average Rating for current record: 7.2
tconst for current record: tt0377091
**********Record Content 31 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30275
Average Rating for current record: 5.9
tconst for current record: tt0417001
**********Record Content 32 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30319
Average Rating for current record: 7.4
tconst for current record: tt1174730
**********Record Content 33 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30326
Average Rating for current record: 7.2
tconst for current record: tt3385524
**********Record Content 34 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30333
Average Rating for current record: 7.8
tconst for current record: tt0880502
**********Record Content 35 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30341
Average Rating for current record: 2.9
tconst for current record: tt0105643
**********Record Content 36 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30354
Average Rating for current record: 7.2
tconst for current record: tt0439662

**********Record Content 37 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30361
Average Rating for current record: 7.2
tconst for current record: tt1629757
**********Record Content 38 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30370
Average Rating for current record: 7.3
tconst for current record: tt1094249
**********Record Content 39 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30376
Average Rating for current record: 7
tconst for current record: tt0482088
**********Record Content 40 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30391
Average Rating for current record: 7.5
tconst for current record: tt0408777
**********Record Content 41 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30395
Average Rating for current record: 7.4
tconst for current record: tt1672723
**********Record Content 42 at address: 0x7fe0e2fa6560 **********
Record address in disk: 0x7fe0e2fa6560
NumVotes for current record: 30402
Average Rating for current record: 8.3
tconst for current record: tt0059026

Summarised Information:
total number of records accessed: 42
Average avgRating of the records: 6.91905
Number of index blocks accesses : 5
Number of data blocks accesses   : 5

Experiment 5 Results:

----------Experiment 5----------
height of B+ Tree is: 5

DONE REMOVING~~~~~~~~

height of B+ Tree is: 5
Content of Parent Node:
****
Node's Disk address: 0x7fe0e302a220
currentKeySize :4
currentPointerSize :5
node type :InternalNode
max Key Size: 13
keyArray: [3591,7911,16859,47952,0,0,0,0,0,0,0,0,0,]

PointerArray:
[0x7fe0e302a090,0x7fe0e3063368,0x7fe0e309c578,0x7fe0e30d5788,0x7fe0e30d5850,0,0,0,0,0,0,0,0,0,]
****
Concent of Child Node:
****
Node's Disk address: 0x7fe0e302a090
currentKeySize :7
currentPointerSize :8
node type :InternalNode
max Key Size: 13
keyArray: [453,901,1349,1797,2245,2693,3141,0,0,0,0,0,0,]
PointerArray:
[0x7fe0e2fcd368,0x7fe0e2fd4640,0x7fe0e2fdb850,0x7fe0e2fe2a60,0x7fe0e2fe9c70,0x7fe0e2ff0e80,0x7fe0e2ff8090,0x7fe0e2fff2a0,0,0,0,0,0,0,]
****
here: 0x7fe0e2ea8bb8
Deleting those movies with the attribute 'numVotes' equal to 1000----------
No. of times that a node is deleted (or two nodes are merged): 1
No. of nodes in updated B+ Tree : 3
Height of updated B+ tree : 5

Block Size: 500B
Experiment 1 Results:
Number of Records : 1070319 Total number of blocks : 44725 Actual size of database : 21335175
Experiment 2 Results:
----------Experiment 2---------- ******BPTREE DETAILS****** Parameter n of B+ Tree (number of keys in Node): 38 Total NodeSize: 1952 nodes height of B+ Tree is: 4 height of B+ Tree is: 4 **Content Of Root Node:** ********** Node's Disk address: 0x7f3dc8623668 currentKeySize :1 currentPointerSize :2 node type :InternalNode max Key Size: 38 keyArray: [8650,0,0,0,0,0,0,0,0,0

[0x7f3dc8623280,0x7f3dc8623474,0,0,0

3nd index block accessed:
****
Node's Disk address: 0x7f3dc8565370
currentKeySize :19
currentPointerSize :20
node type :InternalNode
max Key Size: 38
keyArray:
[404,423,442,461,480,499,518,537,556,575,594,613,632,651,670,689,708,727,746,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,]
PointerArray:
[0x7f3dc855b53c,0x7f3dc855b924,0x7f3dc855bd0c,0x7f3dc855c0f4,0x7f3dc855c4dc,0x7f3dc855
c8c4,0x7f3dc855ccac,0x7f3dc855d094,0x7f3dc855d47c,0x7f3dc855d864,0x7f3dc855dc4c,0x7f3d
c855e034,0x7f3dc855e41c,0x7f3dc855e804,0x7f3dc855ebec,0x7f3dc855efd4,0x7f3dc855f3bc,0x
7f3dc855f7a4,0x7f3dc855fb8c,0x7f3dc8560550,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,]
****
4nd index block accessed:
****
Node's Disk address: 0x7f3dc855ccac
currentKeySize :19
currentPointerSize :20
node type :LEAF
max Key Size: 38
keyArray:
[499,500,501,502,503,504,505,506,507,508,509,510,511,512,513,514,515,516,517,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,0,0,0,0,0,0,]
PointerArray:
[0x7f3dc83997f8,0x7f3dc839a054,0x7f3dc839a8ec,0x7f3dc839b120,0x7f3dc839ba58,0x7f3dc839
c2a0,0x7f3dc839ca70,0x7f3dc839d204,0x7f3dc839da60,0x7f3dc839e3e8,0x7f3dc839eba4,0x7f3
dc839f428,0x7f3dc839fbf8,0x7f3dc83a04e0,0x7f3dc83a0d14,0x7f3dc83a1458,0x7f3dc83a1da4,0
x7f3dc83a25b0,0x7f3dc83a2e34,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0x7f3dc855d094,]
****

Content of data blocks accesed:

**********Record Content 1 at address: 0x7f3dc839a054 **********
Record address in disk: 0x7f3dc839a054
NumVotes for current record: 500
Average Rating for current record: 5.1
tconst for current record: tt0082841
**********Record Content 2 at address: 0x7f3dc839a054 **********
Record address in disk: 0x7f3dc839a054
NumVotes for current record: 500
Average Rating for current record: 4.1
tconst for current record: tt1727519
**********Record Content 3 at address: 0x7f3dc839a054 **********
Record address in disk: 0x7f3dc839a054
NumVotes for current record: 500
Average Rating for current record: 5.5
tconst for current record: tt6038946
**********Record Content 4 at address: 0x7f3dc839a054 **********
Record address in disk: 0x7f3dc839a054
NumVotes for current record: 500
Average Rating for current record: 8
tconst for current record: tt4392820
**********Record Content 5 at address: 0x7f3dc839a054 **********

Record address in disk: 0x7f3dc839a054
NumVotes for current record: 500
Average Rating for current record: 7.1
tconst for current record: tt0082275
**********Record Content 6 at address: 0x7f3dc839a054 **********
Record address in disk: 0x7f3dc839a054
NumVotes for current record: 500
Average Rating for current record: 8
tconst for current record: tt4627488
**********Record Content 7 at address: 0x7f3dc839a054 **********
Record address in disk: 0x7f3dc839a054
NumVotes for current record: 500
Average Rating for current record: 7.4
tconst for current record: tt1632756
**********Record Content 8 at address: 0x7f3dc839a054 **********
Record address in disk: 0x7f3dc839a054
NumVotes for current record: 500
Average Rating for current record: 7.8
tconst for current record: tt1857596
**********Record Content 9 at address: 0x7f3dc839a054 **********
Record address in disk: 0x7f3dc839a054
NumVotes for current record: 500
Average Rating for current record: 7.7
tconst for current record: tt1640740
**********Record Content 10 at address: 0x7f3dc839a054 **********
Record address in disk: 0x7f3dc839a054
NumVotes for current record: 500
Average Rating for current record: 5.3
tconst for current record: tt1684558

Summarised info:(With a block limit of 5)

total number of records accessed: 110
Average avgRating of the records: 6.73182

Number of index blocks accesses : 4
Number of record blocks accesses : 5

Experiment 4 Results:

Content of index blocks accessed:

**********Index Node Content**********
index root Node accessed:
****
Node's Disk address: 0x7f3dc8623668
currentKeySize :1
currentPointerSize :2
node type :InternalNode
max Key Size: 38
keyArray: [8650,0,0,0

2nd index block accessed:
****
Node's Disk address: 0x7f3dc8623474
currentKeySize :26
currentPointerSize :27
node type :InternalNode
max Key Size: 38
keyArray:
[9276,9998,10811,11729,12685,13738,14976,16426,17937,19659,21616,23934,26851,30246,338
12,38357,43416,49838,57181,67581,80192,96442,117031,149235,198248,281245,0,0,0,0,0,0,0,0
,0,0,0,0,]
PointerArray:
[0x7f3dc85c6a08,0x7f3dc85cbc10,0x7f3dc85d0e18,0x7f3dc85d6020,0x7f3dc85db228,0x7f3dc85
e0430,0x7f3dc85e5638,0x7f3dc85ea840,0x7f3dc85efa48,0x7f3dc85f4c50,0x7f3dc85f9e58,0x7f3d
c85ff060,0x7f3dc8604268,0x7f3dc8609470,0x7f3dc860e678,0x7f3dc8613880,0x7f3dc8618a88,0x
7f3dc861dc90,0x7f3dc8622e98,0x7f3dc862867c,0x7f3dc862d884,0x7f3dc8632a8c,0x7f3dc8637c
94,0x7f3dc863ce9c,0x7f3dc86420a4,0x7f3dc86472ac,0x7f3dc86474a0,0,0,0,0,0,0,0,0,0,0,0,0,]
****
3nd index block accessed:
****
Node's Disk address: 0x7f3dc8609470
currentKeySize :19
currentPointerSize :20
node type :InternalNode
max Key Size: 38
keyArray:
[26994,27115,27222,27409,27535,27718,27942,28087,28252,28454,28597,28796,28962,29167,2
9317,29585,29743,29880,30041,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,]
PointerArray:
[0x7f3dc85ff830,0x7f3dc85ffc18,0x7f3dc8600000,0x7f3dc86003e8,0x7f3dc86007d0,0x7f3dc8600
bb8,0x7f3dc8600fa0,0x7f3dc8601388,0x7f3dc8601770,0x7f3dc8601b58,0x7f3dc8601f40,0x7f3dc
8602328,0x7f3dc8602710,0x7f3dc8602af8,0x7f3dc8602ee0,0x7f3dc86032c8,0x7f3dc86036b0,0x
7f3dc8603a98,0x7f3dc8603e80,0x7f3dc8604650,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,]
****
4nd index block accessed:
****
Node's Disk address: 0x7f3dc8603e80
currentKeySize :19
currentPointerSize :20
node type :LEAF
max Key Size: 38
keyArray:
[29880,29882,29900,29910,29919,29949,29954,29956,29959,29962,29974,29975,29978,29982,
29988,29996,30022,30034,30037,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,]
PointerArray:
[0x7f3dc853c3f8,0x7f3dc853c40c,0x7f3dc853c420,0x7f3dc853c434,0x7f3dc853c448,0x7f3dc853
c45c,0x7f3dc853c470,0x7f3dc853c484,0x7f3dc853c4ac,0x7f3dc853c4c0,0x7f3dc853c4d4,0x7f3d
c853c4e8,0x7f3dc853c510,0x7f3dc853c524,0x7f3dc853c538,0x7f3dc853c54c,0x7f3dc853c560,0
x7f3dc853c574,0x7f3dc853c588,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0x7f3dc8604650,]
****

Example of Content of data blocks:

**********Record Content 1 at address: 0x7f3dc853c560 **********
Record address in disk: 0x7f3dc853c560
NumVotes for current record: 30022
Average Rating for current record: 7.7
tconst for current record: tt0054167
**********Record Content 2 at address: 0x7f3dc853c560 **********
Record address in disk: 0x7f3dc853c560
NumVotes for current record: 30034
Average Rating for current record: 7.9
tconst for current record: tt0026778
**********Record Content 3 at address: 0x7f3dc853c560 **********
Record address in disk: 0x7f3dc853c560
NumVotes for current record: 30037
Average Rating for current record: 5.6
tconst for current record: tt0091828
**********Record Content 4 at address: 0x7f3dc853c560 **********
Record address in disk: 0x7f3dc853c560
NumVotes for current record: 30041
Average Rating for current record: 6.8
tconst for current record: tt3361792
**********Record Content 5 at address: 0x7f3dc853c560 **********
Record address in disk: 0x7f3dc853c560
NumVotes for current record: 30049
Average Rating for current record: 6.2
tconst for current record: tt1456941

Summarised info:

total number of records accessed: 122
Average avgRating of the records: 6.82623
Number of index blocks accesses : 4
Number of data blocks accesses   : 5

Experiment 5 Results:

----------Experiment 5----------
height of B+ Tree is: 4

DONE REMOVING~~~~~~~~

height of B+ Tree is: 4
Content of Parent Node:
****
Node's Disk address: 0x7f3dc8623668
currentKeySize :1
currentPointerSize :2
node type :InternalNode
max Key Size: 38
keyArray: [8650,0,0,0

```
0,0,0,0,]
**********
Concent of Child Node:
**********
Node's Disk address: 0x7f3dc8623280
currentKeySize :19
currentPointerSize :20
node type :InternalNode
max Key Size: 38
keyArray:
[385,1145,1525,1905,2285,2665,3045,3427,3811,4202,4595,5005,5429,5874,6356,6866,7419,79
85,9276,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,]
PointerArray:
[0x7f3dc855ff74,0x7f3dc8565370,0x7f3dc856f780,0x7f3dc8574988,0x7f3dc8579b90,0x7f3dc857e
d98,0x7f3dc8583fa0,0x7f3dc85891a8,0x7f3dc858e3b0,0x7f3dc85935b8,0x7f3dc85987c0,0x7f3dc
859d9c8,0x7f3dc85a2bd0,0x7f3dc85a7dd8,0x7f3dc85acfe0,0x7f3dc85b21e8,0x7f3dc85b73f0,0x7
f3dc85bc5f8,0x7f3dc85c1800,0x7f3dc85c6a08,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,]
**********
here: 0x7f3dc843ea8c
Deleting those movies with the attribute 'numVotes' equal to 1000----------
No. of times that a node is deleted (or two nodes are merged): 2
No. of nodes in updated B+ Tree                    : 7
Height of updated B+ tree                          : 4
```

4. Installation guide

Firstly, download GitHub from https://git-scm.com/downloads.

After installing GitHub, run the following commands:

```
git clone https://github.com/kevin-otter-liu/DatabasePrinciplesProject1/tree/main
```

For MacOS and Linux, you can run the program in 2 ways with the following commands
Running the executable:

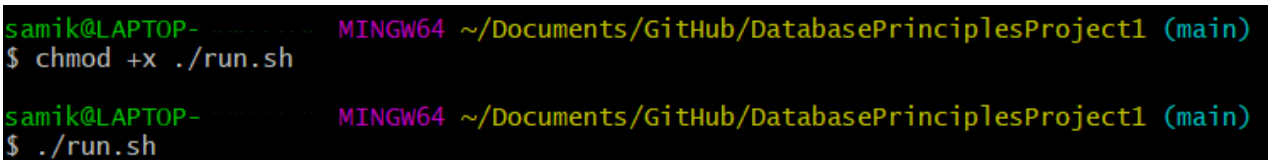
```
chmod +x main
./main
```

compile and run

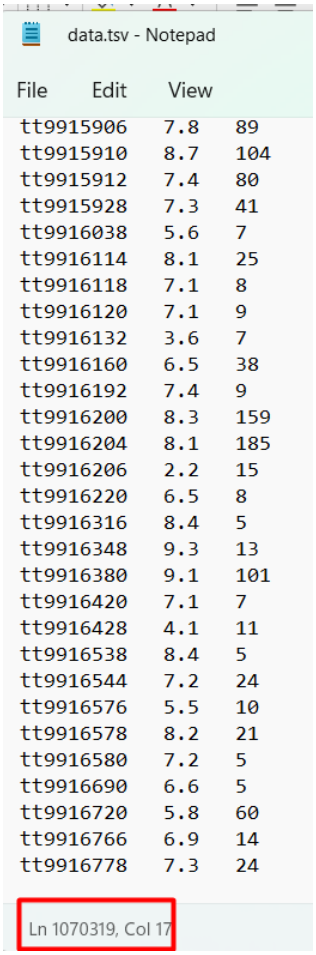
```
chmod +x run_unix.sh
./run_unix.sh
```

For Windows,

1. Follow the following instructions to install the C++ compiler for MINGW64 using MSYS2
    a. https://www.freecodecamp.org/news/how-to-install-c-and-cpp-compiler-on-windows/
2. To directly install MINGW64, follow the following link to install the C++ compiler
    a. https://www.eclipse.org/4diac/documentation/html/installation/minGW.html
3. Thereafter run the following commands in the folder container the codes



5. Appendix:



There's 1070319 lines of data.