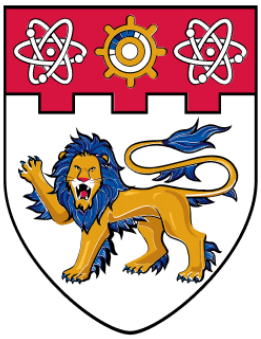


NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

# CZ4031 Database Systems Principles
# Project 1 Group 31 Report


| | |
|---|---|
| WANG CHUHAN | U2022829K |
| Leong Ruo Qing | U1921880A |
| Jerome Chew | U2021304E |
| Karanam Akshit | U2020311E |
| Lau Chen Yi Wynne | U2020016B |

# Content

# 1. Overview

## 1.1 Project Overview

This project is implemented using C++ and consists of the following:

- Storage component, which controls the insertion and deletion of records into blocks.
- B+ tree structure implementation, which supports inserting, searching and deleting operations.

## 1.2 Instructions

### Running on Windows

To install gcc

- Step 1: Install Mingw-w64 via Msys2. https://www.msys2.org
- Step 2: Once installed, a Mingw-w64 environment is opened in a window. Install the full Mingw-w64 toolchain pacman -S --needed base-devel mingw-w64-x86_64-toolchain
- Step 3: Add Mingw-w64 bin folder to your Windows PATH environment variable.
- Step 4: In a new terminal window, to check if gcc has been installed correctly, type gcc --version. If you don't see the expected output or g++ or gdb is not a recognized command, make sure your PATH entry matches the Mingw-w64 binary location where the compiler tools are located.

Run the program

- Step 1 : Clone the repository from: https://github.com/akshitkaranam/CZ4031_Project1
- Step 2 : Open a new terminal window, and navigate to the project directory
- Step 3 : Compile all the relevant files in the src folder by running the following commands:

```
cd src
g++ -std=c++11 -c Address.cpp Record.cpp Block.cpp
Storage.cpp Node.cpp InternalNode.cpp LeafNode.cpp
AddressNode.cpp BPlusTree.cpp
```

- Step 4 : Compile main.cpp file by running the following commands:

```
g++ -std=c++11 main.cpp -o main  Address.o Record.o
Block.o Storage.o Node.o InternalNode.o LeafNode.o
AddressNode.o BPlusTree.o
```

- Step 5 : Run the executable file by running the following command:

```
main
```

## Running on macOS

Run the program

- Step 1 : Clone the repository from: https://github.com/akshitkaranam/CZ4031_Project1
- Step 2 : Open a new terminal window, and navigate to the project directory
- Step 3 : Compile all the relevant files in the src folder by running the following commands

```
cd src
g++ -std=c++11 -c Address.cpp Record.cpp Block.cpp
Storage.cpp Node.cpp InternalNode.cpp LeafNode.cpp
AddressNode.cpp BPlusTree.cpp
```

- Step 4 :  Compile main.cpp file by running the following commands

```
g++ -std=c++11 main.cpp -o main  Address.o Record.o
Block.o Storage.o Node.o InternalNode.o LeafNode.o
AddressNode.o BPlusTree.o
```

- Step 5 : To run the exectuable file by running the following command

```
./main
```

# 2. Storage Component

## 2.1 How each data item is stored as a field

Each field consists of the following data items:

```
class Record {


public:
    bool isDeleted;
    char tconst[10];
    float averageRating;
    int numVotes;
```

- tconst – alphanumeric unique identifier of the title
  For tconst, we can see that the maximum length of a tconst string is 10 characters. Hence, a character array of length 10 is used to store this data so that the field is of fixed-length. When a tconst string has a length of less than 10, the remaining character arrays will be instantiated with null values.

- averageRating – weighted average of all the individual user ratings
  For averageRating, the data provided is a number that has up to 1 decimal place, hence a float is used to store the data.

- numVotes – number of votes the title has received
  For numVotes, the largest value has a length of 7 characters. Since the integer type in C++ can store values from -2147483648 to 2147483647, it is sufficient to use the integer data type to store this data.

- isDeleted – marker to store whether the record has been deleted
  An additional attribute, isDeleted, is a marker to store whether the record has been deleted. This data is stored using the boolean data type, with true indicating that the record has been deleted, and thus be replaced by any incoming record.

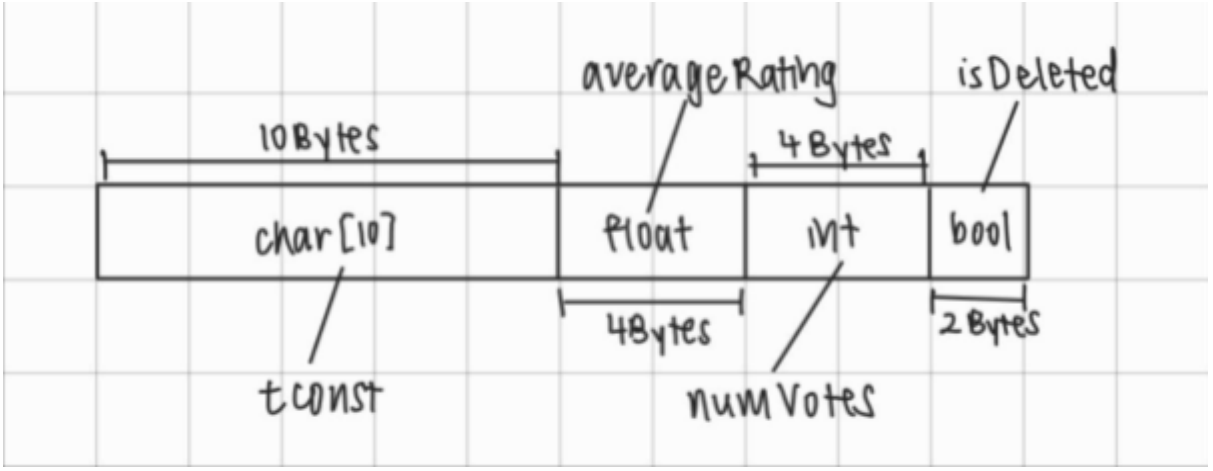
## 2.2 How fields are packed into a record

- tconst
  In C++, the number of bytes for a character variable is 1 byte. As a result, the total number of bytes for a character array of length 10 is 10 bytes.

- averageRating
  In C++, the total number of bytes for a float data type is 4 bytes.

- numVotes
  In C++, the total number of bytes for an integer data type is 4 bytes.

- isDeleted
  In C++, the total number of bytes for a boolean data type is 2 bytes.




10+4+4+2= 20 bytes.

As seen in the above diagram, in total, each record will have a fixed format with a fixed length of 20 bytes. Since the number of bytes per record is fixed, it will be easier to perform insert or delete operations without confusion, although this might lead to some space wasted during storage.

## 2.3 How records are packed into a block

Records are then stored into blocks in a random order (into any available space on any block). Since all records have a fixed length of 20 bytes, it is not necessary to separate the recording using additional techniques. If the block is 200 bytes, the block can hold up to 10 records. If the block is 500 bytes, the block can hold up to 25 records.

```
if (!blocksThatHaveDeletes.empty()) {
    insertedOffset = blocks.at(blocksThatHaveDeletes.at(0)).insertRecord(record);
    return new Address(blocks.at(blocksThatHaveDeletes.at(0)).getNumber(),insertedOffset);
}
```

As stated, each record has a 'isDeleted' attribute to mark whether it has been deleted from the block. Each time a record is to be inserted, the existing blocks will be checked to see if they contain any record that is marked as deleted. Should such a record exist, it means that the space originally taken up by that record is now available for insertion of a new record. The new record will be inserted at that particular address of the block.

```
if(!blocks.empty()){
    if(blocks.at(blocks.size()-1).getAvailableSpace()){
        insertedOffset = blocks.at(blocks.size()-1).insertRecord(record);
        return new Address(blocks.size()-1,insertedOffset);
    }
}
```

Else, record will be inserted at the next available block.

```
if (blocks.size() < maxBlocks) {
    Block newBlock = Block(blockSize,blocks.size()-1);
    blocks.push_back(newBlock);
    blocks.at(blocks.size() - 1).insertRecord(record);
    curBlocks++;
    return new Address(curBlocks - 1, insertedOffset);
} else {
    cout << "Storage is full!" << endl;
    return nullptr;
}
```

If existing blocks are all full (i.e. no records in existing blocks are marked as deleted), the next block is accessed to store records.

Storage is considered to be full when there is no available space.

# 3. B+ Tree

## 3.1. B+ Tree Node

There are two types of nodes in B+ Tree, internal (non-leaf) nodes and leaf nodes. We build a base class (Node class) as well as two derived classes (Internal Node class, Leaf Node class) for B+ Tree nodes.

In this project, we build a B+ tree on the attribute "numVotes" which has duplicate keys. As a result, we design a address list (Address Node class) for each key in leaf node to maintain all addresses of inserted records which has same key value. Since the address node is not a typical B+ tree node, it is not bounded by the block size.

Node class:

```
class Node
{
private:
    vector<int> keys;
```

Internal Node class:

```
class InternalNode : public Node {
private:
    vector<Node *> children;
```

Leaf Node class:

```
class LeafNode : public Node
{

private:
    vector<AddressNode*> addressNodes;
    LeafNode *next{};
```

Address Node class:

```
class AddressNode {

private:
    vector<Address*> addresses;
```

Address class:

```
class Address
{
    int blockNumber;
    int offset;
```

In our design, two derived classes contain the following attributes:
Internal Node:

- vector<int> keys: Each internal node contains a list of keys. In this project, we build a B+ tree on the attribute "numVotes", as a result, we utilize a list of integers to store the key.
- vector<Node *> children: Each internal node contains a list of pointers. We utilize a list of pointers which point to child nodes belongs to the current node.

Leaf Node:

- vector<int> keys: Each leaf node contains a list of keys. In this project, we build a B+ tree on the attribute "numVotes", as a result, we utilize a list of integers to store the key.
- vector<AddressNode*> addressNodes: Each leaf node contains a list of pointers. We utilize a list of pointers to point to the address node with the specific key value.
- Node * nextNode: Each leaf node contains a pointer which points to the next leaf node.

In our design, Address Node class contains the following attribute:

- vector<Address*> addresses: Each address node contains a list of pointers. We utilize a list of pointers to point to inserted records which has same key value.

In our design, Address class contains the following attributes:

- int blockNumber: This attribute refers to the ID of the block.
- int offset: This attribute refers to the offset of the records in a specific block.

## 3.2. B+ Tree

Object reference takes up 8 bytes (64-bit processor) since it is a pointer. One integer variable takes up 4 byte.

In our design, B+ Tree class contains the following attributes:

- int maxKeys: This attribute refers to the maximum number of keys that one node can store. The formula is:

  $$maxKeys = (blockSize - size\_pointer) / (size\_key + size\_pointer)$$

  For block size of **200 B**, each B+ Tree node can store up to 16 keys and 17 pointers.
  For block size of **500 B**, each B+ Tree node can store up to 41 keys and 42 pointers.

- int parentMinKeys: This attribute refers to the minimum number of keys that an internal node can store. The formula is:

  $$floor(maxKeys / 2)$$

  For block size of **200 B**, each B+ Tree node needs to store at least 8 keys.
  For block size of **500 B**, each B+ Tree node needs to store at least 20 keys.

- int leafMinKeys: This attribute refers to the minimum number of keys that a leaf node can store. The formula is:

  $$floor((maxKeys + 1)/ 2)$$

  For block size of **200 B**, each B+ Tree node needs to store at least 8 keys.
  For block size of **500 B**, each B+ Tree node needs to store at least 21 keys.

- Node *root: This attribute refers to the root node of the B+ tree.

- int height: This attribute refers to the height of the B+ tree.

- int nodeCount: This attribute refers to the number of current nodes of the B+ tree.

- int deletedCount: This attribute refers to the node of deleted nodes of the B+ tree.

# 4. Results

## 4.1 Experiment 1

Store the data (which is about IMDb movies) on the disk and report the following statistics:

- the number of blocks
- the size of database (in terms of MB)

| Block Size | Number of blocks | Size of database |
|---|---|---|
| 200 B | 107032 | 21.4064 MB |
| 500 B | 42813 | 21.4065 MB |

For block size of **200 B**,

```
Adding the movies data into storage and creating an index on numVotes.
Number of blocks used : 107032
Size of the database (in MB) : 21.4064
```

For block size of **500 B**,

```
Adding the movies data into storage and creating an index on numVotes.
Number of blocks used : 42813
Size of the database (in MB) : 21.4065
```

## 4.2 Experiment 2

Build a B+ tree on the attribute "numVotes" by inserting the records sequentially and report the following statistics:

- the parameter n of the B+ tree
- the number of nodes of the B+ tree
- the height of the B+ tree, i.e., the number of levels of the B+ tree
- the content of the root node and its 1st child node

| Block Size | Parameter n of the B+ tree | Number of nodes of the B+ tree | Height of the B+ tree | Content of the root node and its 1st child node |
|---|---|---|---|---|
| 200 B | 16 | 1755 | 4 | Refer to figure below |
| 500 B | 41 | 655 | 3 | Refer to figure below |

For block size of **200 B**,

```
Tree statistics:
Max Keys in an index node (n): 16
Number of nodes in B+Tree: 1755
Height of B+Tree: 4
Content of root node: [1544,2985,5254,7328,9132,12582,15380,27444,47273,122013,]
Content of first child node: [132,250,338,478,641,775,905,1092,1214,1308,1438,]
```

For block size of **500 B**,

```
Tree statistics:
Max Keys in an index node (n): 41
Number of nodes in B+Tree: 655
Height of B+Tree: 3
Content of root node: [810,1970,2998,3824,4754,5795,6963,8596,10593,12944,15653,19079,23444
,28454,33701,41156,50539,72884,106475,209225,]
Content of first child node: [26,59,88,120,160,188,222,253,291,316,338,360,395,418,444,484,
506,528,553,575,611,652,676,699,727,754,780,]
```

## 4.3 Experiment 3

Retrieve those movies with the “numVotes” equal to 500 and report the following statistics:

- the number and the content of index nodes the process accesses
- the number and the content of data blocks the process accesses
- the average of “averageRating’s” of the records that are returned

| Block Size | Number of index nodes the process accesses | Number of data blocks the process accesses | Average of “averageRating’s” of the records that are returned |
|---|---|---|---|
| 200 B | 4 | 110 | 6.73182 |
| 500 B | 3 | 110 | 6.73182 |

For the content of index nodes and data blocks the process accesses, refer below:

For block size of **200 B**,

```
Starting Experiment 3
Index Nodes accessed:
1: [1544,2985,5254,7328,9132,12582,15380,27444,47273,122013,]
2: [132,250,338,478,641,775,905,1092,1214,1308,1438,]
3: [487,499,515,528,537,551,562,574,582,591,603,615,630,]
4: [499,500,501,502,503,504,505,506,507,508,509,510,511,512,513,514,]
Total number of Index Nodes accessed: 4
```

```
Data Blocks accessed:
Block Number: 360
Record 1 tconst : tt0013631, averageRating : 6.6, numVotes : 12
Record 2 tconst : tt0013658, averageRating : 6.9, numVotes : 31
Record 3 tconst : tt0013662, averageRating : 6.9, numVotes : 418
Record 4 tconst : tt0013668, averageRating : 6.7, numVotes : 22
Record 5 tconst : tt0013672, averageRating : 6.7, numVotes : 25
Record 6 tconst : tt0013674, averageRating : 7, numVotes : 500
Record 7 tconst : tt0013679, averageRating : 6.9, numVotes : 7
Record 8 tconst : tt0013681, averageRating : 5.6, numVotes : 14
Record 9 tconst : tt0013682, averageRating : 7.5, numVotes : 64
Record 10 tconst : tt0013687, averageRating : 7.1, numVotes : 7

Block Number: 902
Record 1 tconst : tt0024550, averageRating : 6.4, numVotes : 24
Record 2 tconst : tt0024551, averageRating : 2.9, numVotes : 9
Record 3 tconst : tt0024553, averageRating : 5.9, numVotes : 137
Record 4 tconst : tt0024554, averageRating : 5.3, numVotes : 822
Record 5 tconst : tt0024555, averageRating : 5.5, numVotes : 195
Record 6 tconst : tt0024558, averageRating : 6.4, numVotes : 11
Record 7 tconst : tt0024559, averageRating : 6.1, numVotes : 140
Record 8 tconst : tt0024560, averageRating : 6.9, numVotes : 397
Record 9 tconst : tt0024561, averageRating : 6.8, numVotes : 500
Record 10 tconst : tt0024562, averageRating : 5.4, numVotes : 10


Block Number: 1183
Record 1 tconst : tt0028271, averageRating : 6, numVotes : 6
Record 2 tconst : tt0028272, averageRating : 6.1, numVotes : 282
Record 3 tconst : tt0028273, averageRating : 6.2, numVotes : 28
Record 4 tconst : tt0028274, averageRating : 6.6, numVotes : 71
Record 5 tconst : tt0028275, averageRating : 6.5, numVotes : 1246
Record 6 tconst : tt0028276, averageRating : 5.9, numVotes : 30
Record 7 tconst : tt0028277, averageRating : 7.7, numVotes : 500
Record 8 tconst : tt0028278, averageRating : 7.6, numVotes : 5
Record 9 tconst : tt0028279, averageRating : 6.5, numVotes : 131
Record 10 tconst : tt0028280, averageRating : 5.7, numVotes : 35

Block Number: 2278
Record 1 tconst : tt0041946, averageRating : 5.5, numVotes : 39
Record 2 tconst : tt0041947, averageRating : 6.1, numVotes : 517
Record 3 tconst : tt0041948, averageRating : 6.6, numVotes : 902
Record 4 tconst : tt0041949, averageRating : 6.3, numVotes : 355
Record 5 tconst : tt0041951, averageRating : 7.4, numVotes : 53
Record 6 tconst : tt0041952, averageRating : 7.6, numVotes : 690
Record 7 tconst : tt0041953, averageRating : 6.9, numVotes : 469
Record 8 tconst : tt0041954, averageRating : 7.3, numVotes : 2435
Record 9 tconst : tt0041955, averageRating : 6.7, numVotes : 1119
Record 10 tconst : tt0041956, averageRating : 6.5, numVotes : 500


Block Number: 2719
Record 1 tconst : tt0047355, averageRating : 5.9, numVotes : 99
Record 2 tconst : tt0047356, averageRating : 7, numVotes : 10
Record 3 tconst : tt0047357, averageRating : 6.4, numVotes : 27
Record 4 tconst : tt0047358, averageRating : 5.9, numVotes : 224
Record 5 tconst : tt0047359, averageRating : 6.7, numVotes : 20
Record 6 tconst : tt0047360, averageRating : 4.6, numVotes : 12
Record 7 tconst : tt0047361, averageRating : 7.3, numVotes : 500
Record 8 tconst : tt0047362, averageRating : 5.8, numVotes : 5
Record 9 tconst : tt0047363, averageRating : 6.7, numVotes : 30
Record 10 tconst : tt0047364, averageRating : 4.3, numVotes : 11

Number of Data Blocks accessed: 110
Average Rating : 6.73182
```

For block size of **500 B**,

```
Starting Experiment 3
Index Nodes accessed:
1: [810,1970,2998,3824,4754,5795,6963,8596,10593,12944,15653,19079,23444,28454,33701,41156,50539,
72884,106475,209225,]
2: [26,59,88,120,160,188,222,253,291,316,338,360,395,418,444,484,506,528,553,575,611,652,676,699,
727,754,780,]
3: [484,485,486,487,488,489,490,491,492,493,494,495,496,497,498,499,500,501,502,503,504,505,]
Total number of Index Nodes accessed: 3


Data Blocks accessed:
Block Number: 144
Record 1 tconst : tt0013590, averageRating : 5.9, numVotes : 101
Record 2 tconst : tt0013592, averageRating : 6.3, numVotes : 365
Record 3 tconst : tt0013596, averageRating : 3.9, numVotes : 10
Record 4 tconst : tt0013597, averageRating : 5.8, numVotes : 551
Record 5 tconst : tt0013603, averageRating : 4.1, numVotes : 24
Record 6 tconst : tt0013607, averageRating : 5.8, numVotes : 158
Record 7 tconst : tt0013611, averageRating : 6.9, numVotes : 70
Record 8 tconst : tt0013615, averageRating : 3.2, numVotes : 13
Record 9 tconst : tt0013617, averageRating : 5.5, numVotes : 14
Record 10 tconst : tt0013619, averageRating : 6, numVotes : 169
Record 11 tconst : tt0013620, averageRating : 5.6, numVotes : 32
Record 12 tconst : tt0013624, averageRating : 6.5, numVotes : 21
Record 13 tconst : tt0013626, averageRating : 6.7, numVotes : 2031
Record 14 tconst : tt0013627, averageRating : 5.2, numVotes : 10
Record 15 tconst : tt0013629, averageRating : 6.7, numVotes : 25
Record 16 tconst : tt0013631, averageRating : 6.6, numVotes : 12
Record 17 tconst : tt0013658, averageRating : 6.9, numVotes : 31
Record 18 tconst : tt0013662, averageRating : 6.9, numVotes : 418
Record 19 tconst : tt0013668, averageRating : 6.7, numVotes : 22
Record 20 tconst : tt0013672, averageRating : 6.7, numVotes : 25
Record 21 tconst : tt0013674, averageRating : 7, numVotes : 500
Record 22 tconst : tt0013679, averageRating : 6.9, numVotes : 7
Record 23 tconst : tt0013681, averageRating : 5.6, numVotes : 14
Record 24 tconst : tt0013682, averageRating : 7.5, numVotes : 64
Record 25 tconst : tt0013687, averageRating : 7.1, numVotes : 7
```

```
Block Number: 361
Record 1 tconst : tt0024536, averageRating : 4.8, numVotes : 13
Record 2 tconst : tt0024537, averageRating : 6.4, numVotes : 303
Record 3 tconst : tt0024538, averageRating : 6.4, numVotes : 645
Record 4 tconst : tt0024539, averageRating : 6.5, numVotes : 578
Record 5 tconst : tt0024542, averageRating : 1.6, numVotes : 10
Record 6 tconst : tt0024545, averageRating : 4.2, numVotes : 57
Record 7 tconst : tt0024546, averageRating : 6.8, numVotes : 25
Record 8 tconst : tt0024547, averageRating : 5.2, numVotes : 300
Record 9 tconst : tt0024548, averageRating : 6.3, numVotes : 5257
Record 10 tconst : tt0024549, averageRating : 6, numVotes : 361
Record 11 tconst : tt0024550, averageRating : 6.4, numVotes : 24
Record 12 tconst : tt0024551, averageRating : 2.9, numVotes : 9
Record 13 tconst : tt0024553, averageRating : 5.9, numVotes : 137
Record 14 tconst : tt0024554, averageRating : 5.3, numVotes : 822
Record 15 tconst : tt0024555, averageRating : 5.5, numVotes : 195
Record 16 tconst : tt0024558, averageRating : 6.4, numVotes : 11
Record 17 tconst : tt0024559, averageRating : 6.1, numVotes : 140
Record 18 tconst : tt0024560, averageRating : 6.9, numVotes : 397
Record 19 tconst : tt0024561, averageRating : 6.8, numVotes : 500
Record 20 tconst : tt0024562, averageRating : 5.4, numVotes : 10
Record 21 tconst : tt0024563, averageRating : 6, numVotes : 93
Record 22 tconst : tt0024564, averageRating : 6, numVotes : 55
Record 23 tconst : tt0024567, averageRating : 6, numVotes : 8
Record 24 tconst : tt0024568, averageRating : 6.7, numVotes : 25
Record 25 tconst : tt0024569, averageRating : 5.6, numVotes : 142


Block Number: 474
Record 1 tconst : tt0028276, averageRating : 5.9, numVotes : 30
Record 2 tconst : tt0028277, averageRating : 7.7, numVotes : 500
Record 3 tconst : tt0028278, averageRating : 7.6, numVotes : 5
Record 4 tconst : tt0028279, averageRating : 6.5, numVotes : 131
Record 5 tconst : tt0028280, averageRating : 5.7, numVotes : 35
Record 6 tconst : tt0028281, averageRating : 6.4, numVotes : 51
Record 7 tconst : tt0028282, averageRating : 6.5, numVotes : 278
Record 8 tconst : tt0028283, averageRating : 5.2, numVotes : 134
Record 9 tconst : tt0028284, averageRating : 6.1, numVotes : 105
Record 10 tconst : tt0028285, averageRating : 6.2, numVotes : 13
Record 11 tconst : tt0028286, averageRating : 5.9, numVotes : 187
Record 12 tconst : tt0028287, averageRating : 6.9, numVotes : 7
Record 13 tconst : tt0028288, averageRating : 6.6, numVotes : 45
Record 14 tconst : tt0028289, averageRating : 5.5, numVotes : 11
Record 15 tconst : tt0028290, averageRating : 5.9, numVotes : 120
Record 16 tconst : tt0028291, averageRating : 5.9, numVotes : 418
Record 17 tconst : tt0028292, averageRating : 5.6, numVotes : 36
Record 18 tconst : tt0028294, averageRating : 7, numVotes : 173
Record 19 tconst : tt0028296, averageRating : 7.2, numVotes : 192
Record 20 tconst : tt0028297, averageRating : 5.1, numVotes : 8
Record 21 tconst : tt0028298, averageRating : 5, numVotes : 11
Record 22 tconst : tt0028299, averageRating : 6.6, numVotes : 23
Record 23 tconst : tt0028300, averageRating : 6.6, numVotes : 17
Record 24 tconst : tt0028301, averageRating : 4.7, numVotes : 34
Record 25 tconst : tt0028302, averageRating : 7, numVotes : 68
```

```
Block Number: 912
Record 1 tconst : tt0041952, averageRating : 7.6, numVotes : 690
Record 2 tconst : tt0041953, averageRating : 6.9, numVotes : 469
Record 3 tconst : tt0041954, averageRating : 7.3, numVotes : 2435
Record 4 tconst : tt0041955, averageRating : 6.7, numVotes : 1119
Record 5 tconst : tt0041956, averageRating : 6.5, numVotes : 500
Record 6 tconst : tt0041957, averageRating : 7.7, numVotes : 10
Record 7 tconst : tt0041958, averageRating : 7.6, numVotes : 5154
Record 8 tconst : tt0041959, averageRating : 8.1, numVotes : 155251
Record 9 tconst : tt0041961, averageRating : 6, numVotes : 308
Record 10 tconst : tt0041962, averageRating : 6.4, numVotes : 25
Record 11 tconst : tt0041963, averageRating : 6.7, numVotes : 914
Record 12 tconst : tt0041966, averageRating : 6.4, numVotes : 248
Record 13 tconst : tt0041967, averageRating : 6.4, numVotes : 1793
Record 14 tconst : tt0041968, averageRating : 7.3, numVotes : 3571
Record 15 tconst : tt0041969, averageRating : 6.1, numVotes : 216
Record 16 tconst : tt0041971, averageRating : 5.6, numVotes : 143
Record 17 tconst : tt0041974, averageRating : 6.5, numVotes : 396
Record 18 tconst : tt0041975, averageRating : 6.8, numVotes : 395
Record 19 tconst : tt0041976, averageRating : 5.5, numVotes : 76
Record 20 tconst : tt0041977, averageRating : 6.6, numVotes : 19
Record 21 tconst : tt0041978, averageRating : 7.6, numVotes : 1106
Record 22 tconst : tt0041979, averageRating : 6.5, numVotes : 55
Record 23 tconst : tt0041980, averageRating : 5.9, numVotes : 11
Record 24 tconst : tt0041981, averageRating : 5.5, numVotes : 27
Record 25 tconst : tt0041982, averageRating : 6.6, numVotes : 307


Block Number: 1088
Record 1 tconst : tt0047345, averageRating : 5.9, numVotes : 19
Record 2 tconst : tt0047348, averageRating : 5.9, numVotes : 750
Record 3 tconst : tt0047349, averageRating : 6.6, numVotes : 1352
Record 4 tconst : tt0047351, averageRating : 6.5, numVotes : 79
Record 5 tconst : tt0047353, averageRating : 7.1, numVotes : 335
Record 6 tconst : tt0047355, averageRating : 5.9, numVotes : 99
Record 7 tconst : tt0047356, averageRating : 7, numVotes : 10
Record 8 tconst : tt0047357, averageRating : 6.4, numVotes : 27
Record 9 tconst : tt0047358, averageRating : 5.9, numVotes : 224
Record 10 tconst : tt0047359, averageRating : 6.7, numVotes : 20
Record 11 tconst : tt0047360, averageRating : 4.6, numVotes : 12
Record 12 tconst : tt0047361, averageRating : 7.3, numVotes : 500
Record 13 tconst : tt0047362, averageRating : 5.8, numVotes : 5
Record 14 tconst : tt0047363, averageRating : 6.7, numVotes : 30
Record 15 tconst : tt0047364, averageRating : 4.3, numVotes : 11
Record 16 tconst : tt0047365, averageRating : 6.2, numVotes : 2363
Record 17 tconst : tt0047366, averageRating : 6.4, numVotes : 317
Record 18 tconst : tt0047367, averageRating : 5.5, numVotes : 40
Record 19 tconst : tt0047368, averageRating : 5.3, numVotes : 42
Record 20 tconst : tt0047369, averageRating : 4.8, numVotes : 232
Record 21 tconst : tt0047370, averageRating : 6.7, numVotes : 1234
Record 22 tconst : tt0047371, averageRating : 5.4, numVotes : 64
Record 23 tconst : tt0047372, averageRating : 5.7, numVotes : 100
Record 24 tconst : tt0047373, averageRating : 5.8, numVotes : 84
Record 25 tconst : tt0047374, averageRating : 6.9, numVotes : 17

Number of Data Blocks accessed: 110
Average Rating : 6.73182
```

## 4.4 Experiment 4

Retrieve those movies with the attribute“numVotes” from 30,000 to 40,000, both inclusively and report the following statistics:

- the number and the content of index nodes the process accesses
- the number and the content of data blocks the process accesses
- the average of “averageRating’s” of the records that are returned

| Block Size | Number of index nodes the process accesses | Number of data blocks the process accesses | Average of “averageRating’s” of the records that are returned |
|---|---|---|---|
| 200 B | 84 | 953 | 6.72792 |
| 500 B | 38 | 953 | 6.72792 |

For the content of index nodes and data blocks the process accesses, refer below:

For block size of **200 B**,

```
Starting Experiment 4
Getting addressNodes index key in the range: [30000,40000]
Index nodes visited:
1: [1544,2985,5254,7328,9132,12582,15380,27444,47273,122013,]
2: [28659,30177,31254,32246,33357,35490,37547,39659,41859,43396,45674,]
3: [28769,28916,29039,29189,29274,29328,29453,29587,29687,29792,29880,29959,30034,]
4: [29959,29962,29974,29975,29978,29982,29988,29996,30022,]
5: [30034,30037,30041,30049,30053,30056,30078,30081,30085,30090,30136,30144,30149,30158,30168,30175,]
Total number of Index Nodes accessed: 84

Data Blocks accessed:
Block Number: 410
Record 1 tconst : tt0015324, averageRating : 8.2, numVotes : 39659
Record 2 tconst : tt0015329, averageRating : 5.3, numVotes : 156
Record 3 tconst : tt0015331, averageRating : 6.4, numVotes : 55
Record 4 tconst : tt0015339, averageRating : 5.7, numVotes : 71
Record 5 tconst : tt0015342, averageRating : 5.8, numVotes : 74
Record 6 tconst : tt0015343, averageRating : 5.9, numVotes : 182
Record 7 tconst : tt0015347, averageRating : 7.1, numVotes : 17
Record 8 tconst : tt0015349, averageRating : 7.2, numVotes : 58
Record 9 tconst : tt0015353, averageRating : 4.5, numVotes : 11
Record 10 tconst : tt0015355, averageRating : 5.7, numVotes : 11

Block Number: 1066
Record 1 tconst : tt0026776, averageRating : 6.8, numVotes : 73
Record 2 tconst : tt0026777, averageRating : 6.9, numVotes : 15
Record 3 tconst : tt0026778, averageRating : 7.9, numVotes : 30034
Record 4 tconst : tt0026779, averageRating : 5.6, numVotes : 65
Record 5 tconst : tt0026781, averageRating : 6.1, numVotes : 327
Record 6 tconst : tt0026783, averageRating : 6.1, numVotes : 45
Record 7 tconst : tt0026784, averageRating : 6.5, numVotes : 260
Record 8 tconst : tt0026785, averageRating : 5.8, numVotes : 33
Record 9 tconst : tt0026786, averageRating : 5.9, numVotes : 7
Record 10 tconst : tt0026787, averageRating : 6, numVotes : 676
```

```
Block Number: 1235
Record 1 tconst : tt0028950, averageRating : 8.1, numVotes : 33182
Record 2 tconst : tt0028951, averageRating : 5.8, numVotes : 74
Record 3 tconst : tt0028952, averageRating : 6.4, numVotes : 40
Record 4 tconst : tt0028953, averageRating : 6.8, numVotes : 567
Record 5 tconst : tt0028955, averageRating : 6.8, numVotes : 591
Record 6 tconst : tt0028956, averageRating : 6, numVotes : 95
Record 7 tconst : tt0028957, averageRating : 6.5, numVotes : 83
Record 8 tconst : tt0028958, averageRating : 6.3, numVotes : 474
Record 9 tconst : tt0028959, averageRating : 6.7, numVotes : 124
Record 10 tconst : tt0028960, averageRating : 7.9, numVotes : 575

Block Number: 1656
Record 1 tconst : tt0034244, averageRating : 4.8, numVotes : 113
Record 2 tconst : tt0034245, averageRating : 6.8, numVotes : 45
Record 3 tconst : tt0034246, averageRating : 5.3, numVotes : 123
Record 4 tconst : tt0034247, averageRating : 7.4, numVotes : 2710
Record 5 tconst : tt0034248, averageRating : 7.4, numVotes : 31674
Record 6 tconst : tt0034249, averageRating : 6.6, numVotes : 176
Record 7 tconst : tt0034251, averageRating : 7.1, numVotes : 1500
Record 8 tconst : tt0034252, averageRating : 4.5, numVotes : 55
Record 9 tconst : tt0034253, averageRating : 5.5, numVotes : 141
Record 10 tconst : tt0034254, averageRating : 6.1, numVotes : 79

Block Number: 1914
Record 1 tconst : tt0037375, averageRating : 5.3, numVotes : 15
Record 2 tconst : tt0037376, averageRating : 4.4, numVotes : 9
Record 3 tconst : tt0037377, averageRating : 7.4, numVotes : 299
Record 4 tconst : tt0037378, averageRating : 6.4, numVotes : 112
Record 5 tconst : tt0037379, averageRating : 7.2, numVotes : 135
Record 6 tconst : tt0037380, averageRating : 5.6, numVotes : 115
Record 7 tconst : tt0037382, averageRating : 7.8, numVotes : 30254
Record 8 tconst : tt0037383, averageRating : 5.4, numVotes : 14
Record 9 tconst : tt0037384, averageRating : 6.7, numVotes : 677
Record 10 tconst : tt0037385, averageRating : 7.1, numVotes : 263

Number of Data Blocks accessed: 953
Average Rating : 6.72792
```

For block size of **500 B**,

```
Starting Experiment 4
Getting addressNodes index key in the range: [30000,40000]
Index nodes visited:
1: [810,1970,2998,3824,4754,5795,6963,8596,10593,12944,15653,19079,23444,28454,33701,411
56,50539,72884,106475,209225,]
2: [28603,28890,29099,29297,29730,29956,30158,30457,30658,30865,31144,31440,31607,31768,
32028,32195,32470,32644,33044,33357,]
3: [29956,29959,29962,29974,29975,29978,29982,29988,29996,30022,30034,30037,30041,30049,
30053,30056,30078,30081,30085,30090,30136,30144,30149,]
4: [30158,30168,30175,30177,30195,30206,30221,30240,30246,30247,30248,30254,30259,30262,
30275,30319,30326,30333,30341,30354,30361,30370,30376,30391,30395,30402,30418,30423,3043
1,30446,30453,30456,]
5: [30457,30458,30462,30468,30492,30516,30522,30530,30540,30547,30548,30550,30552,30554,
30569,30571,30576,30578,30585,30605,30608,30611,30619,30620,30621,30639,]
Total number of Index Nodes accessed: 38
```

```
Data Blocks accessed:
Block Number: 164
Record 1 tconst : tt0015263, averageRating : 6.2, numVotes : 13
Record 2 tconst : tt0015268, averageRating : 7.1, numVotes : 415
Record 3 tconst : tt0015270, averageRating : 7, numVotes : 14
Record 4 tconst : tt0015273, averageRating : 7, numVotes : 10
Record 5 tconst : tt0015284, averageRating : 6.5, numVotes : 58
Record 6 tconst : tt0015285, averageRating : 6.4, numVotes : 10
Record 7 tconst : tt0015287, averageRating : 6.4, numVotes : 5
Record 8 tconst : tt0015289, averageRating : 6.3, numVotes : 164
Record 9 tconst : tt0015299, averageRating : 7.7, numVotes : 18
Record 10 tconst : tt0015310, averageRating : 7, numVotes : 466
Record 11 tconst : tt0015311, averageRating : 5.4, numVotes : 8
Record 12 tconst : tt0015312, averageRating : 5.5, numVotes : 220
Record 13 tconst : tt0015313, averageRating : 6.7, numVotes : 22
Record 14 tconst : tt0015318, averageRating : 6.7, numVotes : 11
Record 15 tconst : tt0015322, averageRating : 5.1, numVotes : 14
Record 16 tconst : tt0015324, averageRating : 8.2, numVotes : 39659
Record 17 tconst : tt0015329, averageRating : 5.3, numVotes : 156
Record 18 tconst : tt0015331, averageRating : 6.4, numVotes : 55
Record 19 tconst : tt0015339, averageRating : 5.7, numVotes : 71
Record 20 tconst : tt0015342, averageRating : 5.8, numVotes : 74
Record 21 tconst : tt0015343, averageRating : 5.9, numVotes : 182
Record 22 tconst : tt0015347, averageRating : 7.1, numVotes : 17
Record 23 tconst : tt0015349, averageRating : 7.2, numVotes : 58
Record 24 tconst : tt0015353, averageRating : 4.5, numVotes : 11
Record 25 tconst : tt0015355, averageRating : 5.7, numVotes : 11

Block Number: 427
Record 1 tconst : tt0026776, averageRating : 6.8, numVotes : 73
Record 2 tconst : tt0026777, averageRating : 6.9, numVotes : 15
Record 3 tconst : tt0026778, averageRating : 7.9, numVotes : 30034
Record 4 tconst : tt0026779, averageRating : 5.6, numVotes : 65
Record 5 tconst : tt0026781, averageRating : 6.1, numVotes : 327
Record 6 tconst : tt0026783, averageRating : 6.1, numVotes : 45
Record 7 tconst : tt0026784, averageRating : 6.5, numVotes : 260
Record 8 tconst : tt0026785, averageRating : 5.8, numVotes : 33
Record 9 tconst : tt0026786, averageRating : 5.9, numVotes : 7
Record 10 tconst : tt0026787, averageRating : 6, numVotes : 676
Record 11 tconst : tt0026788, averageRating : 5.8, numVotes : 122
Record 12 tconst : tt0026789, averageRating : 6.4, numVotes : 81
Record 13 tconst : tt0026790, averageRating : 4.6, numVotes : 44
Record 14 tconst : tt0026791, averageRating : 5.8, numVotes : 32
Record 15 tconst : tt0026792, averageRating : 6.7, numVotes : 10
Record 16 tconst : tt0026793, averageRating : 6.7, numVotes : 251
Record 17 tconst : tt0026794, averageRating : 4.1, numVotes : 25
Record 18 tconst : tt0026796, averageRating : 7.2, numVotes : 43
Record 19 tconst : tt0026797, averageRating : 5.4, numVotes : 82
Record 20 tconst : tt0026798, averageRating : 5.8, numVotes : 40
Record 21 tconst : tt0026799, averageRating : 6.5, numVotes : 146
Record 22 tconst : tt0026802, averageRating : 5.6, numVotes : 8
Record 23 tconst : tt0026803, averageRating : 4.7, numVotes : 6
Record 24 tconst : tt0026805, averageRating : 6.5, numVotes : 229
Record 25 tconst : tt0026806, averageRating : 6.5, numVotes : 59
```

```
Block Number: 494
Record 1 tconst : tt0028934, averageRating : 5.8, numVotes : 125
Record 2 tconst : tt0028935, averageRating : 6.4, numVotes : 25
Record 3 tconst : tt0028936, averageRating : 6.3, numVotes : 24
Record 4 tconst : tt0028937, averageRating : 7.7, numVotes : 137
Record 5 tconst : tt0028938, averageRating : 6.1, numVotes : 8
Record 6 tconst : tt0028939, averageRating : 6.5, numVotes : 172
Record 7 tconst : tt0028940, averageRating : 6.9, numVotes : 34
Record 8 tconst : tt0028941, averageRating : 5.9, numVotes : 183
Record 9 tconst : tt0028942, averageRating : 5.7, numVotes : 71
Record 10 tconst : tt0028943, averageRating : 5.6, numVotes : 66
Record 11 tconst : tt0028944, averageRating : 7.5, numVotes : 4420
Record 12 tconst : tt0028945, averageRating : 6.7, numVotes : 418
Record 13 tconst : tt0028946, averageRating : 6, numVotes : 48
Record 14 tconst : tt0028947, averageRating : 7.8, numVotes : 448
Record 15 tconst : tt0028949, averageRating : 6.6, numVotes : 21
Record 16 tconst : tt0028950, averageRating : 8.1, numVotes : 33182
Record 17 tconst : tt0028951, averageRating : 5.8, numVotes : 74
Record 18 tconst : tt0028952, averageRating : 6.4, numVotes : 40
Record 19 tconst : tt0028953, averageRating : 6.8, numVotes : 567
Record 20 tconst : tt0028955, averageRating : 6.8, numVotes : 591
Record 21 tconst : tt0028956, averageRating : 6, numVotes : 95
Record 22 tconst : tt0028957, averageRating : 6.5, numVotes : 83
Record 23 tconst : tt0028958, averageRating : 6.3, numVotes : 474
Record 24 tconst : tt0028959, averageRating : 6.7, numVotes : 124
Record 25 tconst : tt0028960, averageRating : 7.9, numVotes : 575


Block Number: 663
Record 1 tconst : tt0034244, averageRating : 4.8, numVotes : 113
Record 2 tconst : tt0034245, averageRating : 6.8, numVotes : 45
Record 3 tconst : tt0034246, averageRating : 5.3, numVotes : 123
Record 4 tconst : tt0034247, averageRating : 7.4, numVotes : 2710
Record 5 tconst : tt0034248, averageRating : 7.4, numVotes : 31674
Record 6 tconst : tt0034249, averageRating : 6.6, numVotes : 176
Record 7 tconst : tt0034251, averageRating : 7.1, numVotes : 1500
Record 8 tconst : tt0034252, averageRating : 4.5, numVotes : 55
Record 9 tconst : tt0034253, averageRating : 5.5, numVotes : 141
Record 10 tconst : tt0034254, averageRating : 6.1, numVotes : 79
Record 11 tconst : tt0034255, averageRating : 6.4, numVotes : 13
Record 12 tconst : tt0034256, averageRating : 6.4, numVotes : 9
Record 13 tconst : tt0034258, averageRating : 4.6, numVotes : 19
Record 14 tconst : tt0034259, averageRating : 6.5, numVotes : 48
Record 15 tconst : tt0034260, averageRating : 7.1, numVotes : 87
Record 16 tconst : tt0034262, averageRating : 6.7, numVotes : 177
Record 17 tconst : tt0034263, averageRating : 5.6, numVotes : 177
Record 18 tconst : tt0034264, averageRating : 6.2, numVotes : 212
Record 19 tconst : tt0034265, averageRating : 6.7, numVotes : 21
Record 20 tconst : tt0034266, averageRating : 6.6, numVotes : 2711
Record 21 tconst : tt0034268, averageRating : 6.6, numVotes : 403
Record 22 tconst : tt0034269, averageRating : 6.7, numVotes : 1098
Record 23 tconst : tt0034270, averageRating : 6.9, numVotes : 15
Record 24 tconst : tt0034272, averageRating : 7.2, numVotes : 3918
Record 25 tconst : tt0034273, averageRating : 6.7, numVotes : 735
```

```
Block Number: 766
Record 1 tconst : tt0037370, averageRating : 7.2, numVotes : 75
Record 2 tconst : tt0037371, averageRating : 5.2, numVotes : 21
Record 3 tconst : tt0037372, averageRating : 5.5, numVotes : 10
Record 4 tconst : tt0037373, averageRating : 5.9, numVotes : 204
Record 5 tconst : tt0037374, averageRating : 6.5, numVotes : 16
Record 6 tconst : tt0037375, averageRating : 5.3, numVotes : 15
Record 7 tconst : tt0037376, averageRating : 4.4, numVotes : 9
Record 8 tconst : tt0037377, averageRating : 7.4, numVotes : 299
Record 9 tconst : tt0037378, averageRating : 6.4, numVotes : 112
Record 10 tconst : tt0037379, averageRating : 7.2, numVotes : 135
Record 11 tconst : tt0037380, averageRating : 5.6, numVotes : 115
Record 12 tconst : tt0037382, averageRating : 7.8, numVotes : 30254
Record 13 tconst : tt0037383, averageRating : 5.4, numVotes : 14
Record 14 tconst : tt0037384, averageRating : 6.7, numVotes : 677
Record 15 tconst : tt0037385, averageRating : 7.1, numVotes : 263
Record 16 tconst : tt0037386, averageRating : 6.5, numVotes : 654
Record 17 tconst : tt0037387, averageRating : 7.1, numVotes : 600
Record 18 tconst : tt0037388, averageRating : 5.7, numVotes : 117
Record 19 tconst : tt0037389, averageRating : 5.8, numVotes : 51
Record 20 tconst : tt0037390, averageRating : 7.1, numVotes : 8
Record 21 tconst : tt0037391, averageRating : 4.9, numVotes : 11
Record 22 tconst : tt0037393, averageRating : 5.8, numVotes : 65
Record 23 tconst : tt0037395, averageRating : 6.6, numVotes : 7
Record 24 tconst : tt0037396, averageRating : 5.8, numVotes : 103
Record 25 tconst : tt0037398, averageRating : 5.7, numVotes : 18

Number of Data Blocks accessed: 953
Average Rating : 6.72792
```

## 4.5 Experiment 5

Delete those movies with the attribute “numVotes” equal to 1,000, update the B+ tree accordingly, and report the following statistics:

- the number of times that a node is deleted (or two nodes are merged) during the process of the updating the B+ tree
- the number nodes of the updated B+ tree
- the height of the updated B+ tree
- the content of the root node and its 1st child node of the updated B+

| Block Size | Number of times that a node is deleted (or two nodes are merged) | Number nodes of the updated B+ tree | Height of the updated B+ tree | Content of the root node and its 1st child node of the updated B+ |
|---|---|---|---|---|
| 200 B | 0 | 1755 | 4 | Refer to figure below |
| 500 B | 0 | 655 | 3 | Refer to figure below |

Since we are using an address list (Address Node class) for each key in leaf node to maintain all addresses of inserted records which has same key value, when we delete those movies with the attribute “numVotes” equal to 1,000, it will only result in the deletion of one key. Hence, the total number of nodes deleted will be 0 as no merging or deletion of nodes are required.

For block size of **200 B**,

```
Starting Experiment 5

Tree statistics:
Max Keys in an index node (n): 16
Number of nodes in B+Tree: 1755
Height of B+Tree: 4
Content of root node: [1544,2985,5254,7328,9132,12582,15380,27444,47273,122013,]
Content of first child node: [132,250,338,478,641,775,905,1092,1214,1308,1438,]
Number of nodes deleted: 0
```

Since the number of nodes deleted is 0, the number of times that a node is deleted (or two nodes are merged) during the process of updating the B+ tree is 0.

For block size of **500 B**,

```
Starting Experiment 5

Tree statistics:
Max Keys in an index node (n): 41
Number of nodes in B+Tree: 655
Height of B+Tree: 3
Content of root node: [810,1970,2998,3824,4754,5795,6963,8596,10593,12944,15653,19079,
23444,28454,33701,41156,50539,72884,106475,209225,]
Content of first child node: [26,59,88,120,160,188,222,253,291,316,338,360,395,418,444
,484,506,528,553,575,611,652,676,699,727,754,780,]
Number of nodes deleted: 0
```

Since the number of nodes deleted is 0, the number of times that a node is deleted (or two nodes are merged) during the process of updating the B+ tree is 0.