Database System (2210)
COSC2406
Assignment 1

Student Name: Ziqing Yan
Student Number: s3749857

Task 1

For building the b+-tree, I used two classes to represent the internal nodes and leaf nodes separately.

Each internal node includes:

1. The flag used to represent the internal node(1 byte).
2. The pointers pointing to child nodes (4 bytes)
3. The keys (8 bytes)

Each leaf node includes:

1. The flag used to represent the leaf node(1 byte)
2. The pointer pointing to the adjacent leaf node on the left(4 bytes)
3. The pointer pointing to the adjacent leaf node on the right(4 bytes)
4. The data entries: keys, page ids and record ids (8 bytes, 4 bytes and 4 bytes)

From the composition of the internal node and leaf node, we can calculate the maximum number of index entries and data entries.

Let M equal to the maximum number of child nodes in an internal node and N equal to the maximum number of data entries. If page size is equal P, we can get

M = (P + 7) / 12 and N = (P - 9) / 16

At the beginning, I read each page from the heap file and parse the birth date as our key. Firstly, I only insert the key into the first leaf node in order. When it is full, the leaf node is split from the middle into two leaf nodes and the two leaf nodes are connected with each other by using pointers. Then, create a new internal node as the current root node and copy up the first key of the right leaf node to the root node. Now, we have a simple tree.

Next, insert the rest of the key into the leaf nodes. This time the key first needs to go through the internal nodes and decide which child node to go to by comparing with the keys in the internal nodes.

As the keys are copied up from leaf nodes to the internal node, the internal node will also be full. After the full internal node is split into two nodes, the first key of the right node will be pushed up to the upper level. If the root node is full, a new node will be created and receive the key from the right node.

After the b+-tree is built, output it as a binary file. What I first write is the page id of the root node in the index file, which is the entry to search for the key. Then, I append the internal nodes with their information to it. Finally, the leaf nodes are added to the binary file.

The instruction of running the program

This task has three .java files, which is btindex.java, InternalNode.java and LeafNode.java. The three files should be together. The heap.xxx file created by task 1 in Assignment is also needed.

The command for running it is showed below:
java btindex -p <pagesize> <heapfile>

For example,
java btindex -p 4096 heap.4096

Task 2

Firstly, read and parse the first 4 bytes in the index file which is the location of the root node. Use the location to find the start point in the index file and parse it in order in terms of the composition of the internal node. When getting the root node, I use the key (the smaller birth date from the command) to compare the keys in the root node and find the corresponding pointer(the location in the index file) pointing to the child node. Find the next node based on the pointer and do the same operation until the next node is a leaf node.

After getting the leaf node, I still use the key(the smaller one) to compare the keys in the leaf node. When finding the first key in the leaf node greater than the smaller key, I start to use the bigger key to compare those keys on its right. If there is a key greater than the bigger key, we stop on it. If we don't find the key greater than the bigger key in the leaf node, use its pointer pointing to its right leaf node to get the right node. We shouldn't stop the operation until there is a key greater than the bigger key.

Whenever I find the key within the range, I use the page id and record id stored in the data entry with the key to find the corresponding page and record in the heap file. Then, parse the record and output it to stdout.

The time of range search without an index is 3556 milliseconds, while that of range search using the index is only 166 milliseconds, which is much faster.

The instruction of running the program

This task has only one file which is btsearch.java. It needs the heap file and index file to run.

The command for running it is showed below:
java btsearch <heapfile> <indexfile> <start date> <end date>

For example,
java btsearch heapfile.4096 index.4096 19700101 19701231

The issue of running btsearch.java on AWS

The program for Task 2 can run and output the result on my computer, but is stuck on an AWS instance. It only stuck there and didn't give me an error message and result.

Below is the result of the program running on my computer:

```
Norm Magnusson     1960-03-20      NULL    1960-03-20      NULL    NULL    NULL    NULL    http://en.wikipedia.org/wiki/Special:FilePath/The_Imposition_of_Order_Upon_Nature
Tore Brunborg      1960-03-20      {Norway|Sør-Trøndelag|Trondheim}        1960-03-20      NULL    Jazz    Saxophone       NULL    NULL    36779442        Jazz saxophonists
Benoît Duteurtre           1960-03-20      {Normandy|Seine-Maritime|Sainte-Adresse}        1960-03-20      NULL    NULL    NULL    NULL    NULL    8453083 French novelist a
Robert Sweet (musician) 1960-03-21         NULL    1960-03-21      NULL    {Heavy metal music|Nu metal|Glam metal|Hard rock|Christian metal}       NULL    NULL    NULL    6
Dominic Miller     1960-03-21      Buenos Aires    1960-03-21      NULL    {Jazz|Pop music|Classical music}        NULL    NULL    http://commons.wikimedia.org/wiki/Special
Owen Blundell      1960-03-22      NULL    1960-03-22      NULL    Country music   NULL    NULL    NULL    19028018        Australian singer
Paul Kater         1960-03-22      NULL    1960-03-22      NULL    NULL    NULL    Netherlands     NULL    37465782        Writer
Ragini Trivedi     1960-03-22      Kanpur  1960-03-22      NULL    Hindustani classical music      {Jal tarang|Sitar|Vichitra veena}       NULL    http://en.wikipedia.org/w
Bojan Pečar        1960-03-22      {Socialist Federal Republic of Yugoslavia|Belgrade}     1960-03-22      NULL    {Rock music|New Wave music in Yugoslavia}       {Bass gui
Vasily Shumov      1960-03-23      {Moscow|Russia|Jacksonville Florida|Florida|United States}      1960-03-23      NULL    {Electronic music|New wave music|Avant-garde|Spre
Espen Beranek Holm         1960-03-23      NULL    1960-03-23      NULL    Pop music       Multi-instrumentalist   NULL    NULL    4427826 Norwegian musician and comedian
George Fox (singer)        1960-03-23      NULL    1960-03-23      NULL    Country music   NULL    NULL    NULL    10081114        Canadian singer
Nena       1960-03-24      {Germany|North Rhine-Westphalia|Hagen}  1960-03-24      NULL    {New wave music|Pop music|Punk rock|Synthpop|Neue Deutsche Welle}        Singing N
Grayson Perry      1960-03-24      Chelmsford      1960-03-24      Fine art        NULL    NULL    NULL    http://en.wikipedia.org/wiki/Special:FilePath/Grayson_perry_portr
Steve Norman       1960-03-25      {England|London}        1960-03-25      NULL    {New wave music|Synthpop|Pop rock}      {Guitar|Percussion instrument|Saxophone}        N
William H S McIntyre       1960-03-25      {Scotland|Linlithgow}   1960-03-25      NULL    NULL    United Kingdom  NULL    42360506        Scottish novelist
Chad Wackerman     1960-03-25      {California|Long Beach California|United States}         1960-03-25      NULL    {Jazz|Rock music|Jazz fusion}   {Drum kit|Percussion inst
Idy Chan           1960-03-25      Hong Kong       1960-03-25      NULL    NULL    NULL    NULL    NULL    1558100 Hong Kong actor
Ingor Ánte Áilo Gaup       1960-03-25      {Norway|Kautokeino}     1960-03-25      NULL    Joik    Joik    NULL    http://commons.wikimedia.org/wiki/Special:FilePath/Inga-N
Linda Sue Park     1960-03-25      Urbana Illinois 1960-03-25      NULL    Young-adult fiction     NULL    Americans       http://commons.wikimedia.org/wiki/Special:FilePat
Nelson Rangell     1960-03-26      NULL    1960-03-26      NULL    Smooth jazz     Piccolo NULL    http://commons.wikimedia.org/wiki/Special:FilePath/Nelson_Rangell.jpg?wid
Renato Russo       1960-03-27      {Brazil|Rio de Janeiro} 1960-03-27      NULL    {Punk rock|Alternative rock|Folk rock|Post-punk}        {Bass guitar|Guitar|Keyboard inst
Victor Bailey (musician)           1960-03-27      Philadelphia    1960-03-27      NULL    {Jazz|Pop music|Jazz fusion}    {Bass guitar|Double bass|Fender Musical Instrumen
Johannes Kerkorrel         1960-03-27      {Johannesburg|South Africa}     1960-03-27      NULL    NULL    NULL    NULL    3073402 South African singer-songwriter
Tony Palacios      1960-03-28      NULL    1960-03-28      NULL    Christian rock  {Guitar|Singing}        NULL    http://en.wikipedia.org/wiki/Special:FilePath/Tony7.jpg?v
Éric-Emmanuel Schmitt      1960-03-28      {Rhône|Sainte-Foy-lès-Lyon}     1960-03-28      NULL    NULL    NULL    Belgium http://commons.wikimedia.org/wiki/Special:FilePat
Jordan S. Gruber   1960-03-28      {Queens|New York}       1960-03-28      NULL    NULL    NULL    United States   http://en.wikipedia.org/wiki/Special:FilePath/Jor
Stephen Street     1960-03-29      NULL    1960-03-29      NULL    {Pop music|Rock music}  NULL    NULL    NULL    558058  British record producer
Jo Nesbø   1960-03-29      Oslo    1960-03-29      NULL    NULL    NULL    Norway  http://commons.wikimedia.org/wiki/Special:FilePath/Jo_Nesbo.jpg?width=300       9
Barry A. Vann      1960-03-30      NULL    1960-03-30      NULL    NULL    NULL    NULL    http://commons.wikimedia.org/wiki/Special:FilePath/Barry_Promo_Photo.jpg?width=36
Popa Chubby        1960-03-31      {The Bronx|New York City|United States} 1960-03-31      NULL    Electric blues  {Drum kit|Guitar|Human voice}   NULL    http://commons.wi
Linda Chorney      1960-03-31      NULL    1960-03-31      NULL    NULL    NULL    NULL    34035159        Singer/songwriter
Shanna McCullough          1960-04-01      NULL    1960-04-01      NULL    NULL    NULL    NULL    http://commons.wikimedia.org/wiki/Special:FilePath/Shanna_McCullough_0101
Frieda Hughes      1960-04-01      {England|London}        1960-04-01      {Poetry|Painting}       NULL    NULL    British people  NULL    2604898 British artist
```

The number of milliseconds to do this query:166 milliseconds