COP 6726: DATABASE SYSTEM IMPLEMENTATION Assignment 1: Heap File Implementation

Group Info:

Name UFID

Manishkumar Chopra 1796-7121
 Neel Manish Rami 7712-3151

I. INTRODUCTION

In this assignment, we have implemented the DBFile class. The job of the DBFile class within the database system is simply to store and retrieve records from the disk. We have also written unit test cases using google test unit testing library for every function which we have implemented in DBFile Class. We have also recorded the execution time of each query and also recorded the time taken to scan all the records.

II. ASSUMPTIONS

- A. gcc (or clang) must be installed together with "bison" and "flex" on the machine on which the project will be tested.
- B. Google Test Unit Testing Library should also be installed and compiled on the machine on which the project will be tested.
- C. Google Test Unit Testing Library is not installed on CISE Thunder Server so the gTest command won't run on thunder.

III. FILE STRUCTURE

- A. bin/: Contains object(.o) files.
- B. src/: Contains .cc and .h files.
- C. table_files/: Contains .tbl files which were generated using tpch data generator.
- D. gtest/: Google Test Unit testing library.
- E. catalog: Catalog file for the schema of .tbl files
- F. MakeFile: Make File.

IV. CODE EXPLANATION

Private Data Structures used in DBFile Class:-

- 1. File* myFile: DBFile Class is built on top of File Class.
- **2. Record* currentRecord:** This variable points to the current record of the DBFile.
- 3. Page* myPage: DBFile Class is built on top of Page Class.
- **4. bool pageWritten:** This variable indicates whether the current page has been written to the DBFile or not.
- **5. Off_t currentPageIndex:** This variable specifies the current page index.

Public Functions used in DBFile Class:-

1. void Add (Record &addMe);

This function is used to add records to the file. In the case of the unordered heap file that has been implemented in this assignment, this function simply adds the new record to the end of the file.

2. int Create (char *name, fType myType, void *startup);

This function is used to create the file.

3. int Open (char *name);

This function is used to open a file that has already been created and also being closed.

4. int Close();

This function simply closes the file.

void Load (Schema &mySchema, char *loadMe);

This function bulk loads the DBFile instance from a text file, appending new data to it using the SuckNextRecord function from Record.h.

6. int GetNext (Record &fetchMe);

This function simply gets the next record from the file and returns it to the user.

7. int GetNext (Record &fetchMe, CNF &applyMe, Record &literal);

This function accepts a selection predicate and returns the next record in the file which is accepted by the selection predicate.

8. void MoveFirst ()

This function forces some pointer to correspond to the first record in the file.

V. COMMANDS TO RUN THE PROGRAM

To compile the given test driver, type make test.out

To run the test driver, type

./test.out

and then follow the on-screen instructions. For scan and filter operation, write any CNF which has been given in the result section below. After writing the CNF, press ENTER and then press Ctrl+D.

The thunder server at cise.ufl.edu doesn't have Google Test installed so you might have to download it.

To compile Google Test, type make gTest

To run the unit tests, type cd bin ./gTestProj1

The unit tests should run and you should be able to see the test results.

VI. RESULTS

We ran our program on 2 sets of data which was generated by tpch data generator. One set of data is 1G and other is 10M.

A. 1G Data

We verified our results with the results which have been provided by the professor in the assignment folder.

Query 1

(r_name = 'EUROPE')

Query 2

(r_name = 'MIDDLE EAST') AND (r_regionkey > 1)

(n_regionkey = 3) AND (n_nationkey > 10) AND (n_name > 'JAPAN')

```
select test:

1. load file
2. scan
3. scan & filter
3

select table:

1. nation
2. region
3. customer
4. part
5. partsupp
6. orders
7. lineitem
1

Filter with CNF for : nation
Enter CNF predicate (when done press ctrl-D):
(n regionkey = 3) AND (n nationkey > 10) AND (n name > 'JAPAN')
nationkey: [19], n name: [ROMANIA], n_regionkey: [3], n_comment: [ular asymptotes are about the furious multipliers. express dependencies nag above the ire lically ironic account]
nationkey: [23], n_name: [ROMSIA], n_regionkey: [3], n_comment: [requests against the platelets use never according to the quickly regular pint]
nationkey: [23], n_name: [UNITED KINGDOM], n_regionkey: [3], n_comment: [eans boost carefully special requests. accounts are. carefull]
elected 3 Records.
uery Time 0.000122seconds.
hunder:1000
```

Query 4

(c_nationkey = 23) AND (c_mktsegment = 'FURNITURE') AND (c_acctbal > 7023.99) AND (c_acctbal < 7110.83)

```
select test:
1. load file
2. scan
3. scan & filter
3. scan s. scan & filter
3. scan & filter
3. scan & filter
3. scan & filter
3. scan & filter
4. nation
2. region
3. sustomer
4. partump
6. orders
7. limeitem
7. limeitem
8. c. c. scan & filter
9. AND (c. mateginent)
8. partump
8. c. orders
7. limeitem
8. c. c. scan & filter
9. scan & filt
```

(p_brand = 'Brand#13') AND (p_retailprice > 500.00) AND (p_retailprice < 930.00) AND (p_size > 28) AND (p_size < 1000000)

```
select test:

1. load file
2. scan
3. scan 4 filter
3
3. scan 4 filter
3. scan 6 filter
4. part
5. partaup
6. orders
7. lineitem
7. lineitem
8. scan 7. lineitem
8. scan 8. sca
```

Query 6

(ps_supplycost > 999.98)

```
select table:

1. nation
2. region
3. customer
4. part
5. partsupp
6. orders
7. lineites
Filter with CMF for: partsupp
Enter CMF predicate (when does press ctrl-D):
(page 12. lineites)
page partsus; (12.13), ps. suppkeys; (37.14), ps. availably; (15070), ps. supplycost: (309.99), ps. comment: [ironic, special deposits, carefully final deposits haggle
ps. partsus; (12.13), ps. suppkeys; (37.14), ps. availably; (1970), ps. supplycost: (399.99), ps. comment: [uriously regular requests haggle, slow foxes at the furiously final accounts hang slyly fluffilly express instructions, permanent, final deposits are sometimes after the blithely unusual packages, fil
ps. partsky; (19079), ps. suppkey; (1821), ps. availably; (1460), ps. supplycost: (1909), ps. comment: [y final requests. final platelets cajele carefully, blith
ely ironic packages wake after the blithely bold requests. sl
ps. partsky; (18241), ps. suppkey; (1824), ps. availably; (1828), ps. supplycost: (1909), ps. comment: [y final requests. requests alongside of the pendi
ps. partsky; (18241), ps. suppkey; (1824), ps. availably; (1828), ps. supplycost: (1909), ps. comment: [uriously regular requests. requests alongside of the pendi
ps. partsky; (1904), ps. suppkey; (1824), ps. availably; (1909), ps. supplycost: (1909), ps. comment: [uriously regular requests. requests. express as aliast the slyly final Tiroisas. blithely regular instructions maintain up the furiously sly final Tiroisas. blithely regular instructions maintain up the furiously sly final Tiroisas. blithely regular instructions maintain up the furiously sly final Tiroisas. blithely regular instructions maintain up the furiously special theol
ps. partsky; (1908), ps. suppkey; (1909), ps. supplycost: (1900), ps. comment: [ans kindle along the blithely even excuses. slyly careful forces are blithely, the doiloites dazzle ironically, ideas sleep among the asymptotes. package slowe the instructions was
ps. partsky; (190830), ps. suppkey; (1908), ps. supplycost: (1900), ps. comment: [ully about the fur
```

(ps_availqty < 10) AND (ps_supplycost > 100.00) AND (ps_suppkey < 300)

```
select table:

1. nation
2. region
3. customer
4. part
5. partsupp
6. orders
7. limeltem
8. partsupp
Enter CNF predicate (when done press ctl-D):
(ps availqty <10) AND (ps supplycost > 100.00) AND (ps supplycost > 200.00)
ps partkey: [7656], ps suppkey: [197], ps availqty: [7], ps supplycost: [226.55], ps comment: [ove the even requests. special deposits sleep. epitaphs boost slowly. final acl
ps partkey: [7556], ps suppkey: [102], ps_availqty: [7], ps_supplycost: [72.6], ps_comment: [ove the even requests. special deposits sleep. epitaphs boost slowly. final acl
ps_partkey: [25577], ps_suppkey: [102], ps_availqty: [1], ps_supplycost: [167.33], ps_comment: [sc. carefully express frays sleep slyly. pending, regular deposits]
ps_partkey: [25573], ps_suppkey: [16], ps_availqty: [2], ps_supplycost: [624.24], ps_comment: [ackages. slyly pending accounts are furiously blithe instructions. oxpress,
partkey: [50757], ps_suppkey: [275], ps_availqty: [2], ps_supplycost: [275.99], ps_comment: [sc. carefully express frays sleep slyly. pending, regular deposits]
ps_partkey: [50757], ps_suppkey: [275], ps_availqty: [3], ps_supplycost: [275.99], ps_comment: [sc. carefully express frays sleep slyly. pending, regular deposits
ps_partkey: [50757], ps_suppkey: [275], ps_availqty: [3], ps_supplycost: [275.99], ps_comment: [sc. carefully express frays sleep slyly. pending, regular deposits
ps_partkey: [50050], ps_suppkey: [275], ps_availqty: [3], ps_supplycost: [275.99], ps_comment: [sc. carefully express frays sleep slyly. ps_partkey: [50050], ps_suppkey: [275], ps_availqty: [3], ps_supplycost: [275.99], ps_comment: [sc. carefully express frays sleep slyly. ps_partkey: [50050], ps_suppkey: [275], ps_availqty: [3], ps_supplycost: [275.99], ps_comment: [sc. carefully express sleep slyly slient deposits are bilthely bold decoys detect furiously infoined psendencies. even, requiar theology.

ps_partkey: [50050], ps_suppkey: [275], ps_availqty: [3], ps_supplycost: [275.99], ps_comment: [sc. carefully even vaters are quickly]

ps_partke
```

Query 8

(o_orderpriority = '1-URGENT') AND (o_orderstatus = '0') AND (o_shippriority = 0) AND (o_totalprice > 1015.68) AND (o_totalprice < 1051.89)

```
load file
2. scan
3. scan & filter
4. part
5. partsupp
6. orders
7. linetem
6. orders
7. linetem
6. orders
7. linetem
7. linetem
7. linetem
8. scan & filter
8. scan & filter
8. scan & filter
9. scan & filter
9.
```

(l_shipdate > '1994-01-01') AND (l_shipdate < '1994-01-04') AND (l_discount > 0.04) AND (l_discount < 0.07) AND (l_quantity = 4.00)

```
labipmode: [RAIL], l_comment: [sual dinos] orderkey: [13761d, l_suppkey: [307], l_inenumber: [1], l_quantity: [4], l_extendedprice: [7255.04], l_discount: [0.05], l_tax: [0.0], l_returnflag: [A], l_linestatus: [F], l_shipdate: [1994-01-03], l_commidate: [1993-12-27], l_eceiptdate: [1994-01-04], l_shipinstruct: [NORE], l_shipmate: [1904.01-04], l_comment: [fully fluffily seen] orderkey: [1750141], l_pattkey: [1900], l_shippate: [1994-01-02], l_commidate: [1994-01-03], l_receiptdate: [1994-01-09], l_shipinstruct: [DELIVER IN PERSO [1, l_shipmate: [1994-01-02], l_commidate: [1994-01-13], l_receiptdate: [1994-01-09], l_shipinstruct: [DELIVER IN PERSO [1, l_stream.com [1, l_shipmate: [1994-01-12], l_commidate: [1994-01-13], l_receiptdate: [1994-01-09], l_shipinstruct: [COLLECT COD), l_shipmote: [1924-01-0], l_commidate: [1994-01-15], l_receiptdate: [1994-01-2], l_discount: [0.05], l_tax: [0.0], l_receiptdate: [1994-01-2], l_commidate: [1994-01-2], l_shipinstruct: [1914-12], l_shipinstruct: [1914-12], l_shipinstruct: [1914-12], l_shipinstruct: [1914-12], l_shipinstruct: [19
```

Query 10

(l_orderkey > 100) AND (l_orderkey < 1000) AND (l_partkey > 100) AND (l_partkey < 5000) AND (l_shipmode = 'AIR') AND (l_linestatus = 'F') AND (l_tax < 0.07)

```
* IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location: catalog
tpch files dir: /cise/tmp/dbi_sp11/DATA/1G/
heap files dir: /cise/tmp/dbi_sp11/DATA/1G/
heap files dir: /cise/tmp/dbi_sp11/DATA/1G/
heap files dir: /cise/tmp/dbi_sp11/DATA/1G/

select test:

1. load file
2. scan
3. scan & filter
3

select table:
1. nation
2. region
3. counter
4. partsupp
6. orders
7. lineitem
7

Filter with CNF for: lineitem
Enter CNF predicate (when done press ctrl-D):
(1. orderkey > 100) AND (1. orderkey < 100) AND (1. partkey > 100) AND (1. partkey < 500) AND (1. shipmode = "AIR") AND (1. linestatus = "F") AND (1. tax
0.07)
orderkey: [130], l_partkey: [1739], l_suppkey: [4240], l_linenumber: [2], l_quantity: [48], l_extendedprice: [78755], l_discount: [0.03], l_tax: [0.02], l_eturnflag: [R], l_linestatus: [F], l_shipdate: [1992-07-01], l_commitdate: [1992-07-12], l_receiptdate: [1992-07-24], l_shipinstruct: [NONE], l_shipmode: [AIR], l_comment: [litely alongside of the regul
orderkey: [194], l_partkey: [2594], l_suppkey: [5095], l_linenumber: [1], l_quantity: [17], l_extendedprice: [25442], l_discount: [0.05], l_tax: [0.04], l_eturnflag: [R], l_linestatus: [F], l_shipdate: [1992-05-24], l_commitdate: [1992-05-22], l_receiptdate: [1992-05-30], l_shipinstruct: [COLLECT COD], l_shipm
de: [AIR], l_comment: [regular deposi]
elected 2 Records.
uery Time 3.76799seconds.
```

We also ran our results on 10M data and verified those results by writing SQL queries in MariaDB.

Google Test Results:

```
neel@nrami:~/Desktop/DBI/P1$ cd bin
neel@nrami:~/Desktop/DBI/P1/bin$ ./gTestProj1
[======] Running 9 tests from 7 test cases.
[-----] Global test environment set-up.
             2 tests from OpenTest
             OpenTest.SubTest1
        2 tests from OpenTest (1 ms total)
             2 tests from CloseTest
             CloseTest.SubTest1
           CloseTest.SubTest1 (0 ms)
            CloseTest.SubTest2
             CloseTest.SubTest2 (0 ms)
        2 tests from CloseTest (0 ms total)
        ---] 1 test from LoadTest
             LoadTest.SubTest1
           LoadTest.SubTest1 (0 ms)
         --] 1 test from LoadTest (0 ms total)
             1 test from AddTest
            AddTest.SubTest1
        OK ] AddTest.SubTest1 (1 ms)
        ---] 1 test from AddTest (1 ms total)
     ----- 1 test from MoveFirstTest
           MoveFirstTest.SubTest1
        OK ] MoveFirstTest.SubTest1 (0 ms)
       ----] 1 test from MoveFirstTest (0 ms total)
       ----] 1 test from GetNext1Test
             GetNext1Test.SubTest1
             GetNext1Test.SubTest1 (0 ms)
           1 test from GetNext1Test (0 ms total)
             1 test from GetNext2Test
             GetNext2Test.SubTest1
        OK GetNext2Test.SubTest1 (0 ms)
       ----] 1 test from GetNext2Test (0 ms total)
     -----] Global test environment tear-down
  ======= 9 tests from 7 test cases ran. (2 ms total)
   PASSED ] 9 tests.
```