COP 6726: DATABASE SYSTEM IMPLEMENTATION Assignment 2.2: Sorted File Implementation

Group Info:

Name UFID

Manishkumar Chopra 1796-7121
 Neel Manish Rami 7712-3151

I. INTRODUCTION

In this assignment, we have implemented the Sorted File. The job of the Sorted File class within the database system is to store in a sorted order wherein the sort order is provided by the user 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.

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* mySFFile: Sorted File Class is built on top of File Class.
- 2. Page* mySFPage: Sorted File Class is built on top of Page Class.
- 3. Pipe* mySFInputPipe: This variable is used to insert records into the Input Pipe.
- **4. Pipe* mySFInputPipe:** This variable is used to remove records from the Output Pipe.
- **5. Mode myMode:** This variable is an enum which has 2 values. One is Read and other is Write
- 6. Sortinfo* mySFSortinfo:
- **7. BigQ* mySFBigQ:** An instance of BigQ which was implemented during the previous project.
- 8. OrderMaker* mySFOrderMaker:
- **9. bool queryFlag:** This variable is used when 2nd version of GetNext is called repeatedly.
- **10. bool sortOrderFlag:** This variable is used when 2nd version of GetNext is called repeatedly.

Public Functions used in DBFile Class:-

void Add (Record &addMe);

This function is used to add records to the file. In the case of the sorted file that has been implemented in this assignment, this function first sees whether it is read or write mode. If it is in read mode, mySFBiQ variable is re-initialized and the record is then inserted into the Input Pipe.

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

This function is used to create the file. Also along with creating the file, we are also creating metadata file for each .tbl file. MetaData File description is written in the code 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. But before closing the file, merge operation is performed.

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

This function bulk loads the records from the file into the Input Record. After all the records have been inserted into Input Pipe, they are removed from the Output Pipe and then they are inserted into the page and these pages are added into the Sorted File.

int GetNext (Record &fetchMe);

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

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.

9. void Merge ()

This function merges records from input pipe with the records from the sorted file. We also first change its mode to read and shutdown the pipe. Then, we merge records from the pipe and sorted 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 gTest21

To run the unit tests, type cd bin ./gTestProj21

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

VI. RESULTS

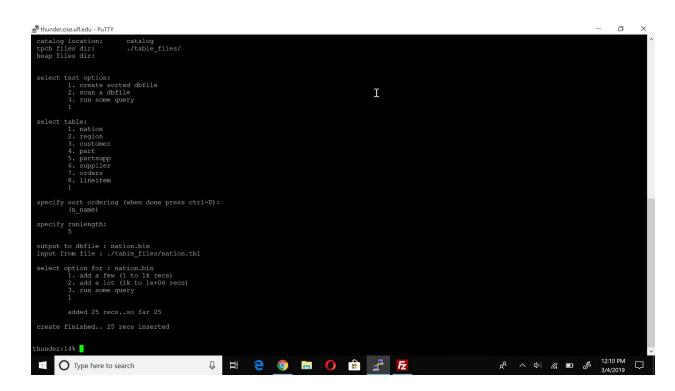
We ran our program on 1 set of data which was generated by tpch data generator. That set of data is 10M.

A. 10M Data

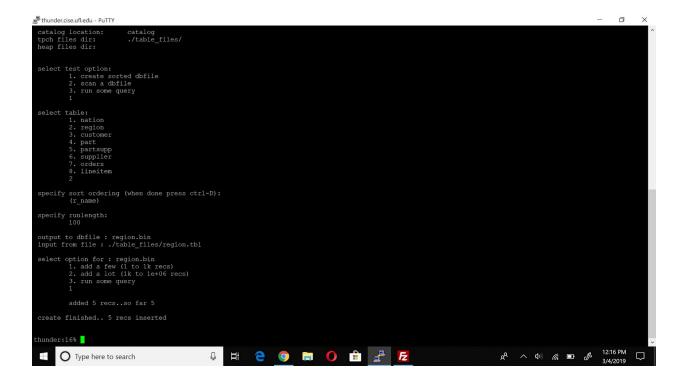
We verified those results by inserting this 10M data into MariaDB and writing SQL queries in MariaDB.

Query Type A: Creating Sorted File

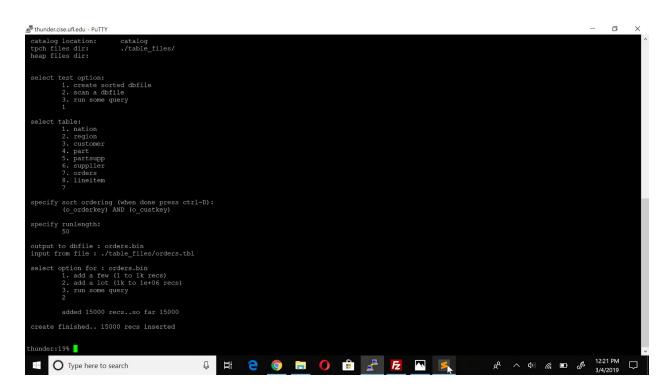
1. nation.tbl



2. region.tbl



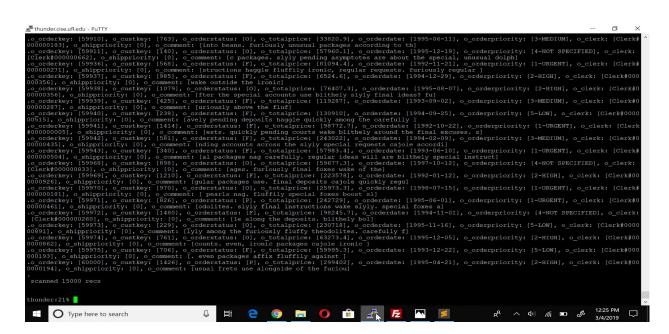
3. orders.tbl



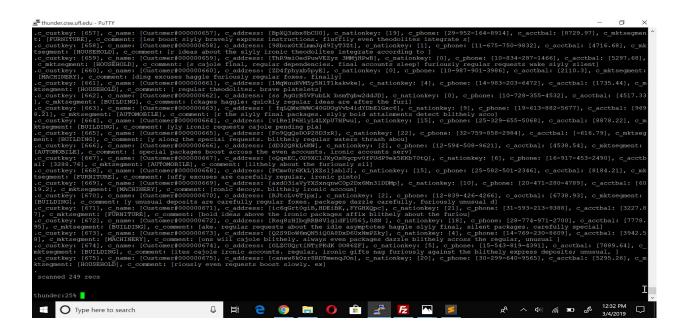
Query Type B (Scanning a Sorted File)

1. nation.tbl

2. orders.tbl



3. Customer.tbl



Query Type C (Extracting Records based on user provided CNF)

Query 1 (r_name = 'EUROPE')

```
thunder:17% ./test.out
 ** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
 catalog location:
tpch files dir:
heap files dir:
                          catalog
                           ./table_files/
 select test option:
          1. create sorted dbfile
          2. scan a dbfile
          3. run some query
 select table:
          1. nation
          2. region
          customer
          4. part
          5. partsupp
          6. supplier
          orders
          8. lineitem
 enter CNF predicate (when done press ctrl-D):
         (r_name = 'EUROPE')
r_regionkey: [3], r_name: [EUROPE], r_comment: [ly final courts cajole furiously final excuse]
 query over region.bin returned 1 recs
thunder:18%
```

Query 2 (r_name = 'MIDDLE EAST') AND (r_regionkey > 1)

```
thunder:4% ./test.out

** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location: catalog
tpch files dir: ./table_files/
heap files dir:

select test option:

1. create sorted dbfile
2. scan a dbfile
3. run some query
3

select table:

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

enter CMF predicate (when done press ctrl-0):
    (r_name = 'MIDDLE EAST') AND (r_regionkey> 1)
    r_regionkey: [4], r_name: [MIDDLE EAST], r_comment: [uickly special accounts cajole carefully blithely close requests. carefully final asymptotes haggle furiousl]

query over region.bin returned 1 recs
```

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

```
thunder:7% ./test.out

** IMPORTANT: NAKE SURE THE IMPORMATION BELON IS CORRECT **
catalog location: catalog
tpch files dir: ./table_files/
heap files dir: ./table_files/
heap files dir:

1. create sorted dbfile
2. scan a dbfile
3. run sone query
3

select table:

1. nation
2. region
3. customer
4. part
5. partsupp
6. suppler
7. orders
8. lineten
1
1 lineten
2 lineten
1 lineten
3 lineten
1 lineten
1 lineten
1 lineten
2 lineten
3 lineten
1 lineten
3 lineten
1 lineten
4 lineten
5 lineten
1 lineten
6 lineten
7 linetenkey: [3], n_name: [ROMANIA], n_regionkey: [3], n_comment: [ular asymptotes are about the furious multipliers. express dependencies nag above the tronically ironic accoun
tj
n_nationkey: [23], n_name: [ROMANIA], n_regionkey: [3], n_comment: [eans boost carefully special requests. accounts are. carefull]
query over nation.bin returned 3 recs
```

Query 4 (c_nationkey = 23) AND (c_mktsegment = 'FURNITURE') AND (c_acctbal > 1000.00) AND (c_acctbal < 8000.00)

```
** INDECRIAT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location: catalog
toch files dir: //table_files/
heap files dir: //table_files/
heap files dir: //table_files/
heap files dir:

1. create sorted doffile
2. sean a doffile
3. run some query
3
celect teather:
1. nation
2. region
3. catalog
6. castomer
6. persump
6. supplier
7. orders
8. Intelten
3. detect (when done press ctrl-D):
C.nationkey = 23 MD (c_nktsegment = FERNITURE!) AND (c_acctbal > 1000.00) AND (c_acctbal < 8000.00)
6. supplier
7. orders
8. Intelten
3. detect (united and the sea of the se
```

Query 5 (p_brand = 'Brand#13') AND (p_retailprice > 500.00) AND (p_retailprice < 930.00) AND (p_size > 28) AND (p_size < 1000000)

Query 6 (ps_supplycost > 999.00)

```
thunder:18% ./test.aut

** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location: catalog
tpch files dir: ./table_files/
heap files dir: ./table_files/
heap files dir: ./table_files/
heap files dir: ./table_files/
heap files dir: ./table_files/
select test option:

select test option:

1. create sorted dofile
2. can a dofile
3. run some query
4. port
5. partsupp
6. supplier
7. orders
8. thetten
8. thetten
9. paper files (files), ps_suppley: [91], ps_avallqty: [5383], ps_supplycost: [999.36], ps_comment: [Instructions. slyly even accounts across the bold ideas hag]
ps_partkey: [167], ps_suppley: [97], ps_avallqty: [97], ps_avallqty: [97], ps_comment: [furlously pending pinto beans. special, bold instructions wake regularly. blitth]
ps_partkey: [167], ps_suppley: [17], ps_avallqty: [273], ps_supplycost: [999.64], ps_comment: [lyv_alongly pending pinto beans. special, bold instructions wake regularly. blitth]
ps_partkey: [167], ps_suppley: [17], ps_avallqty: [273], ps_supplycost: [999.97], ps_comment: [lyv_alongly pending pinto beans. special, bold instructions wake regularly. blitth]
by tronic excuses, quickly regular ideas poach. ideas haggle across the car]
ps_partkey: [167], ps_suppley: [17], ps_avallqty: [167], ps_supplycost: [999.92], ps_comment: [run supply the quickly special deposits, quickly regular platelets use slyly thin
ps_partkey: [187], ps_supplye: [81], ps_supplycost: [999.92], ps_comment: [run supply series [197], accounted the forest paper platelets use slyly thin
ps_partkey: [187], ps_supplye: [81], ps_supplycost: [999.93], ps_comment: [run supply special deposits, quickly regular platelets use slyly thin
ps_partkey: [187], ps_supplye: [81], ps_supplycost: [999.93], ps_comment: [run supply special deposits and printure slanges do fire furiously dog
ps_partkey: [187], ps_supplye: [81], ps_supplycost: [999.93], ps_comment: [run supply special deposits and printurely allonged of the furiou
```

Query 7 (ps_availqty < 10) AND (ps_supplycost > 100.00) AND (ps_suppkey < 300)

```
thunder:19% ./test.out

** IMPORTANT: MAK SURE THE INFORMATION BELOW IS CORRECT **
catalog location: catalog
tph files dir: ./table_files/
heap files dir:

1. create sorted dbfile
2. can abfile
3. run some query
3
select test option:
2. region
3. customer
4. part
5. partsupp
6. supplier
7. orders
8. linetien
Gs. availty < 109 .AND (ps_supplycost : 100.00) AND (ps_supplycost: [102.28], ps_comment: [ regular excuses, quickly ironic courts are bustly, theodolites haggle carefully slyl
y final accounts. regular, special asymptotes haggle along the furfust]
sp_partkey: [747], ps_suppley: [78], ps_availty: [8], ps_supplycost: [68.73], ps_comment: [ express packages try to use above the even excuses. furtously regular deposits above the sile
1]
sp_partkey: [744], ps_suppley: [78], ps_availty: [8], ps_supplycost: [786.9, ps_comment: [ [ sepress packages try to use above the even excuses. furtously regular deposits above the sile
1]
sp_partkey: [747], ps_suppley: [78], ps_availty: [8], ps_supplycost: [786.9, ps_comment: [ [ sepress packages try to use above the even excuses. furtously regular deposits above the sile
1]
sp_partkey: [747], ps_suppley: [81], ps_savailty: [8], ps_supplycost: [786.9, ps_comment: [ [ sepress packages try to use above the even excuses. furtously regular deposits packages wake. final asymptotes ar]
sp_partkey: [747], ps_suppley: [81], ps_savailty: [8], ps_supplycost: [98.62], ps_comment: [ [ sepress packages try to use above the even excuses. furtously regular deposits pseudously regular deposits capide slighting theory and supplycost: [88], ps_comment: [ sepress packages try to use above the even excuses. furtously regular deposits pseudously regular deposits pseu
```

Query 8

(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)

```
thunder:25% /test.out

** IMBORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location:
c
```

```
Lorderey (288). Larstwey (280). Lasupskey: (341). Litemember: (3). Laguarity: (290). Lettendeprice: (4952.2). Lahlocount: (50.2). Lasu (6). Lorderitas: (1939-86-25). Lahlocount: (1939-86-25). Lahlocount: (1939-86-26). Landocount: (1939-86-26). Landocou
```

Google Test Results:

```
neel@nrami:~/Desktop/DBI/Projects/Project2.1/bin$ ./gTestProj21
            Running 7 tests from 7 test cases.
Global test environment set-up.
            1 test from SFCreateTest
            SFCreateTest.SubTest1
       OK ] SFCreateTest.SubTest1 (0 ms)
----] 1 test from SFCreateTest (1 ms total)
            1 test from SFOpenTest
           SFOpenTest.SubTest1
       OK ] SFOpenTest.SubTest1 (34 ms)
       1 test from SFOpenTest (34 ms total)
       ----] 1 test from SFMoveFirstTest
           SFMoveFirstTest.SubTest1
       OK ] SFMoveFirstTest.SubTest1 (3 ms)
       ____] 1 test from SFMoveFirstTest (3 ms total)
       ----] 1 test from SFAddTest
           SFAddTest.SubTest1
       OK ] SFAddTest.SubTest1 (2 ms)
        ---] 1 test from SFAddTest (2 ms total)
            1 test from SFLoadTest
            SFLoadTest.SubTest1
       OK | SFLoadTest.SubTest1 (2 ms)
        --- 1 test from SFLoadTest (2 ms total)
       ----] 1 test from GetNext1Test
            GetNext1Test.SubTest1
       OK GetNext1Test.SubTest1 (2 ms)
        --- 1 test from GetNext1Test (2 ms total)
   -----] 1 test from GetNext2Test
       ----] 1 test from GetNext2Test (2 ms total)
    -----] Global test environment tear-down
 PASSED ] 7 tests from 7 test cases ran. (47 ms total)
neel@nrami:~/Desktop/DBI/Projects/Project2.1/bin$
```