PROJECT REPORT

IMPLEMENTING A SORTED FILE

MILESTONE 2:

EXTENDING THE DBFILE CLASS

Name: Gloria Gilbert Nazareth
UFID: 8221-8035
GatorLink: gnazareth@ufl.edu

Steps to compile and execute:

1. make ./test.out : this command builds the program and creates the required object files
2. ./test.out: this will execute the program, and ask for user input in order to perform required functionality.
3. make ./gtest.out : this command will build the program and create the required object files
4. ./runGtest.sh : this will execute the gtest. We are executing it through a script file, as the yy_parse() function used to test sorting and priority queue requires user input.

Description of the Implemented Function:

The SortedFile class has the following functions:

1. SortedFile ()
   This is the constructor of the class. It initializes the required variables like pageCount, pageNumber and runLength. It also initializes the objects namely Page and File.

2. int Create (char *f_path, fType f_type, void *startup)
   This function creates the sorted file. Before creating the sorted file, it creates a meta file and writes the type of the file i.e sorted, runlength and orderMaker object in it so that it can be used later during file open.

3. int Open(char *f_path):
   This function opens the file in the f_path. It first opens the meta file, reads the runlength and the orderMaker object and assigns it accordingly. It then opens the sorted file and sets the mode to read.

4. void MoveFirst ()
   This function sets the mode to read and the pageNumber to 0(zero) i.e first page of the file.

5. int Close()
   This function sets the mode to read and closes the sorted file.

6. void *producer (void *arg)
   This function is the producer utility which is used to insert records in the input pipe required for sorting. It takes the structure producerUtil as parameter which contains the file path, sortedFile, ordermaker. Once the records are added to it, it shuts down itself.

7. void *consumer (void *arg)
   This function also takes the producerUtil as input. It is responsible to write the sorted records on to the sorted file after sorting.

8. void SortFileAgain()
   This function is responsible for sorting the already sorted and the newly added unsorted records together. It does so, by using p_threads and the BigQ.

9. void SwitchMode(Mode newMode)
   This function performs the necessary action while switching the mode from read to write and vice-versa.

10. void Add (Record &rec)
    This function switches the mode to write and adds one record at a time to file for sorting.

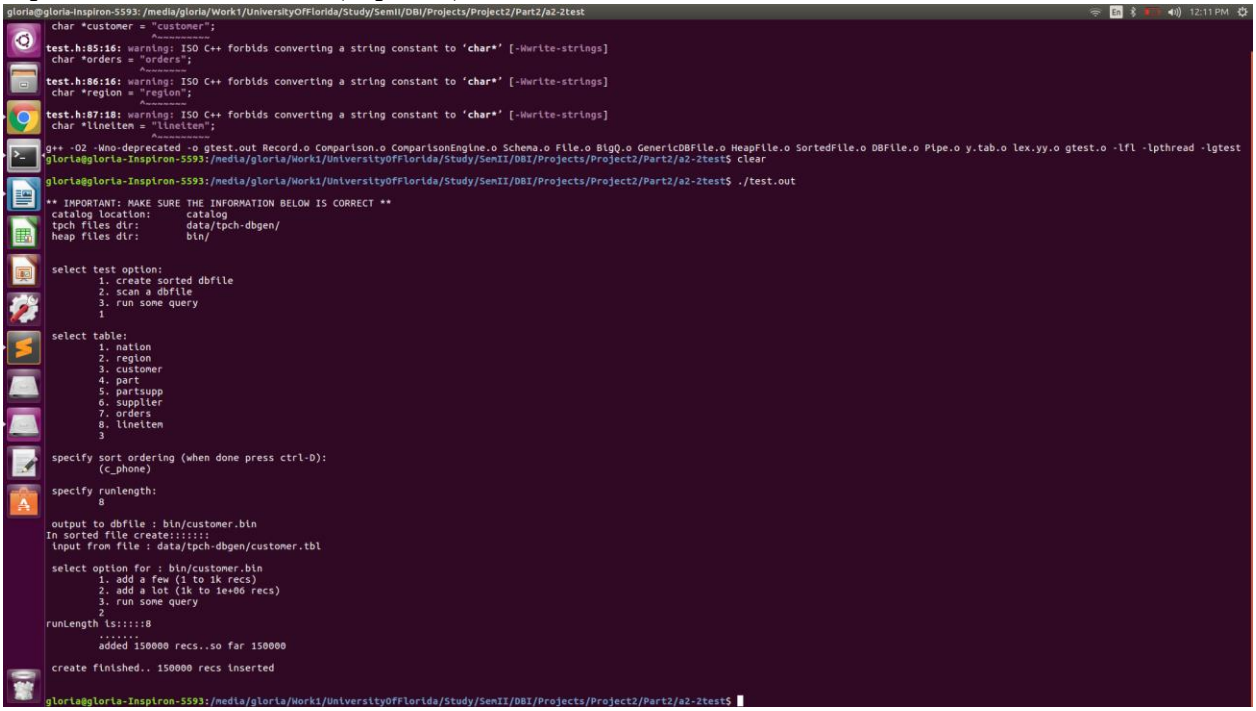
11. void Load (Schema &schema, char *fpath)
    This function calls the producer thread by giving it the fpath. The producer thread when called through this function loads all the data on to the file, adds it in the pipe and sorts it. It uses the BigQ for sorting and the output pipe to write records on to the file after sorting.

12. int GetNext (Record &fetchme)
    This function returns a single record from the sorted file by first switching the mode to read.

13. int GetNext (Record &fetchme, CNF &cnf, Record &literal)
    This function returns the records from the file that match the given CNF. As the file is sorted, fetching records is done efficiently.

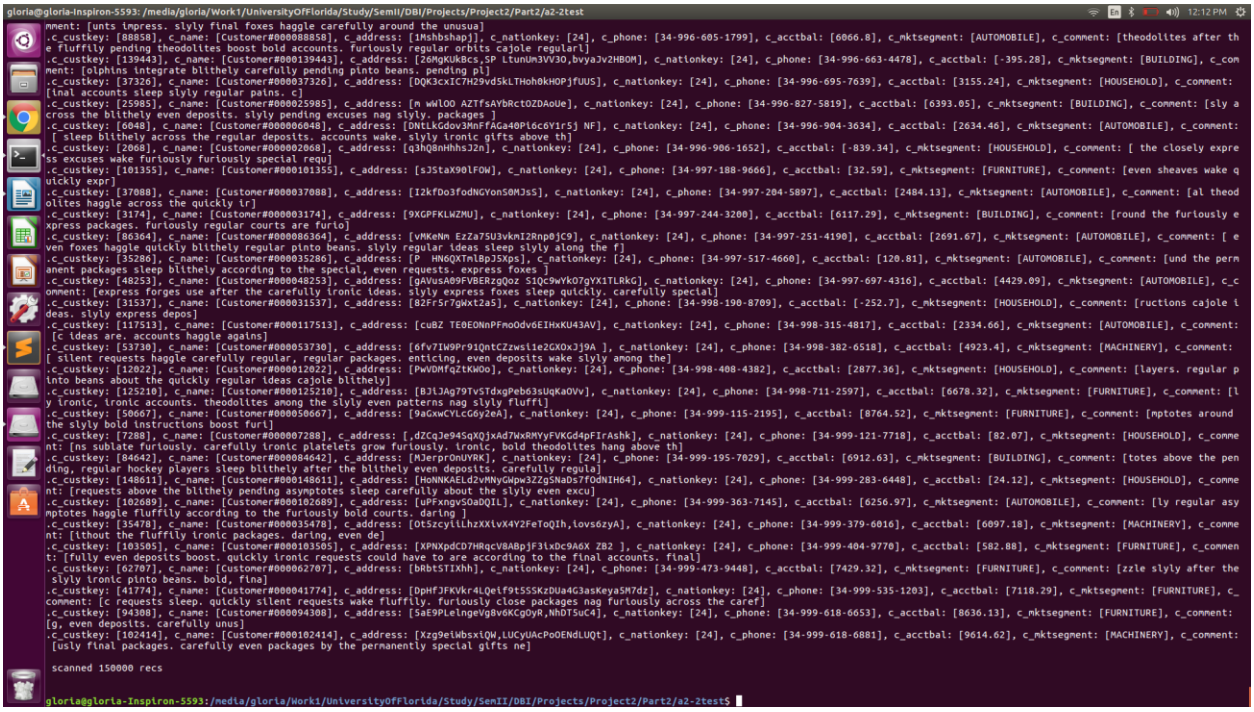
Output on 1GB data:

1. Inputs are as follows: 1 3 (c_phone) 8 2

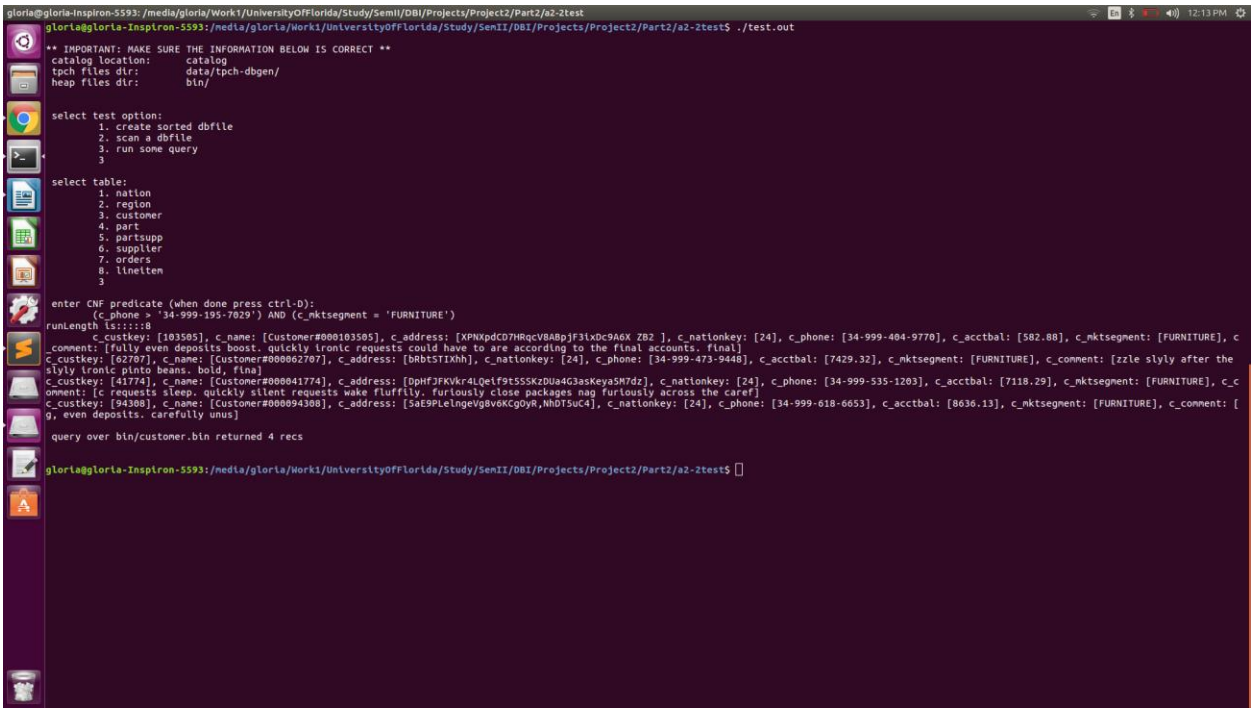


2. Inputs are as follows: 2 3

```
gloria@gloria-Inspiron-5593:/media/gloria/Work1/UniversityOfFlorida/Study/SemII/DBI/Projects/Project2/Part2/a2-2test$ ./test.out

** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
 catalog location:      catalog
 tpch files dir:        data/tpch-dbgen/
 heap files dir:        bin/


 select test option:
        1. create sorted dbfile
        2. scan a dbfile
        3. run some query
        2

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



3. Inputs are as follows: 3 3 (c_phone > '34-999-195-7029') AND (c_mktsegment = 'FURNITURE')



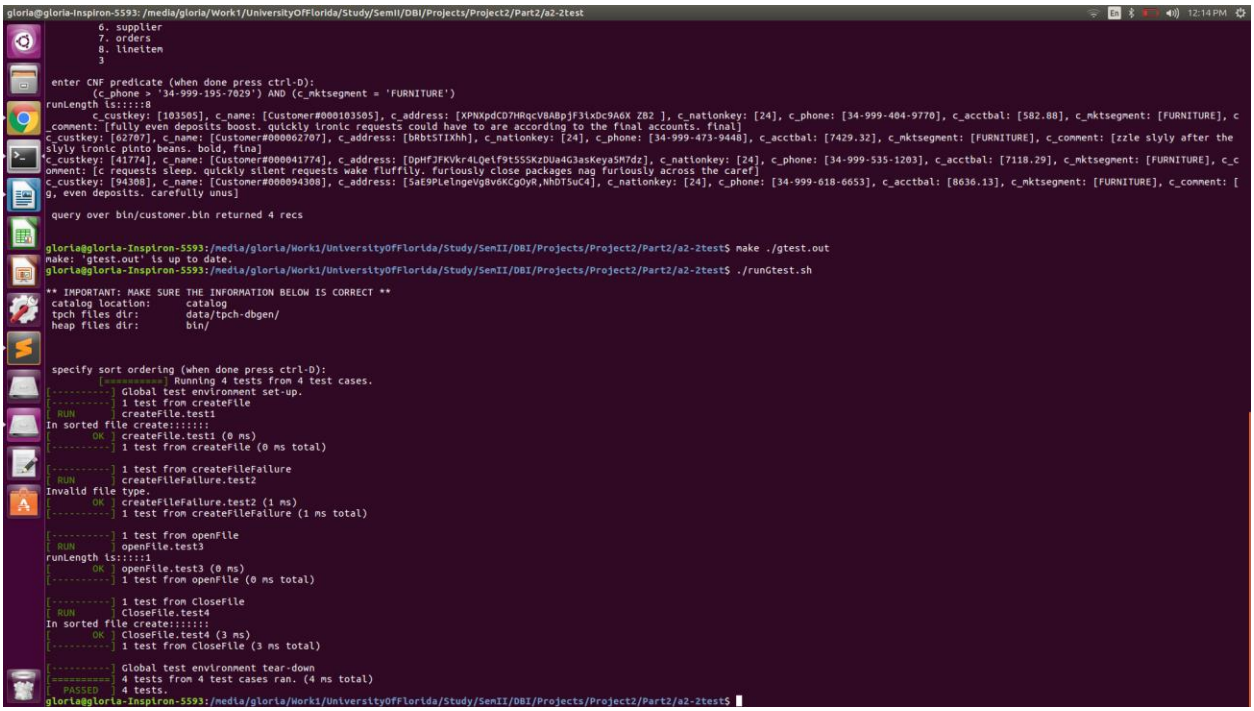
Gtest Results:

1. TEST(createFile,test1)
   This function tests if the create function of the sorted file works fine. It creates a sorted file using the relation region, f_type as sorted and an ordemaker formed on (r_regionkey)

2. TEST(createFileFailure,test2)
   This function illustrates the failure of file creation. It sends the f_type as txt which is invalid.

3. TEST(openFile, test3)
   This function tests open function of the sorted file. It returns 1 if the file has opened properly after reading the meta file.

4. TEST(CloseFile, test4)
   This function checks the functionality of the close function. It closes the file after sorting all the records successfully by switching the mode to read.


gloria@gloria-Inspiron-5593: /media/gloria/Work1/UniversityOfFlorida/Study/SemII/DBI/Projects/Project2/Part2/a2-2test
6. supplier
7. orders
8. lineitem
3
enter CNF predicate (when done press ctrl-D):
(c_phone = '34-999-195-7029') AND (c_mktsegment = 'FURNITURE')
runLength is:::::8
c_custkey: [103505], c_name: [Customer#000103505], c_address: [XPNXpdCD7HRqcV8ABpjF3lxDc9A6X ZB2], c_nationkey: [24], c_phone: [34-999-404-9770], c_acctbal: [582.88], c_mktsegment: [FURNITURE], c_comment: [fully even deposits boost. quickly ironic requests could have to are according to the final accounts. final]
c_custkey: [62707], c_name: [Customer#000062707], c_address: [bRbtSTIXhh], c_nationkey: [24], c_phone: [34-999-473-9448], c_acctbal: [7429.32], c_mktsegment: [FURNITURE], c_comment: [zzle slyly after the slyly ironic pinto beans. bold, fina]
c_custkey: [41774], c_name: [Customer#000041774], c_address: [DpHfJFKVkr4LQeif9t5SSKzDUa4G3asKeyaSM7dz], c_nationkey: [24], c_phone: [34-999-535-1203], c_acctbal: [7118.29], c_mktsegment: [FURNITURE], c_comment: [c requests sleep. quickly silent requests wake fluffily. furiously close packages nag furiously across the caref]
c_custkey: [94308], c_name: [Customer#000094308], c_address: [5aE9PLelngeVg8v6KCgOyR,NhDT5uC4], c_nationkey: [24], c_phone: [34-999-618-6653], c_acctbal: [8636.13], c_mktsegment: [FURNITURE], c_comment: [g, even deposits. carefully unus]
query over bin/customer.bin returned 4 recs
gloria@gloria-Inspiron-5593:/media/gloria/Work1/UniversityOfFlorida/Study/SemII/DBI/Projects/Project2/Part2/a2-2test$ make ./gtest.out
make: 'gtest.out' is up to date.
gloria@gloria-Inspiron-5593:/media/gloria/Work1/UniversityOfFlorida/Study/SemII/DBI/Projects/Project2/Part2/a2-2test$ ./runGtest.sh
** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location: catalog
tpch files dir: data/tpch-dbgen/
heap files dir: bin/
specify sort ordering (when done press ctrl-D):
[==========] Running 4 tests from 4 test cases.
[----------] Global test environment set-up.
[----------] 1 test from createFile
[RUN] createFile.test1
In sorted file create:::::::
[OK] createFile.test1 (0 ms)
[----------] 1 test from createFile (0 ms total)
[----------] 1 test from createFileFailure
[RUN] createFileFailure.test2
Invalid file type.
[OK] createFileFailure.test2 (1 ms)
[----------] 1 test from createFileFailure (1 ms total)
[----------] 1 test from openFile
[RUN] openFile.test3
runLength is:::::1
[OK] openFile.test3 (0 ms)
[----------] 1 test from openFile (0 ms total)
[----------] 1 test from CloseFile
[RUN] CloseFile.test4
In sorted file create:::::::
[OK] CloseFile.test4 (3 ms)
[----------] 1 test from CloseFile (3 ms total)
[----------] Global test environment tear-down
[==========] 4 tests from 4 test cases ran. (4 ms total)
[PASSED] 4 tests.
gloria@gloria-Inspiron-5593:/media/gloria/Work1/UniversityOfFlorida/Study/SemII/DBI/Projects/Project2/Part2/a2-2test$