

Report Assignment 2 Part 2: Implementing a Sorted File

Group Members:

1. Dhiraj Mahesh Paryani, UFID: 1692 1261
2. Yogesh Laxman, UFID: 9451 2517

Instructions for Execution/ Code compilation and run tests:

I. Extracting folders:

- a. Extract the contents of the folder. DhirajMaheshParyani_YogeshLaxman_p2-2.zip
- b. Open the terminal and navigate to the extracted folder.

II. Update test.h:

Update catalog_path, dbfile_dir, & tpch_dir paths in test.h.

III. Running tests:

- A. Execute the following commands. Please note that the tpch files should be generated using the dbgen program before execution. Please change the tpch, dbfile output and catalog directories if necessary, the test.h file.
 - a. `$ make clean`
 - b. `$ make test.out`
 - c. `$./test.out`
- B. This would give a menu-based interface the performs the following three options that can test the code:
 - a. create sorted dbfile
 - b. scan a dbfile
 - c. run some query
- C. Provide inputs for the menu-based interface by specifying the relation, run length, & sort order as required.

Brief explanation of implementation:

As mentioned in the project requirements, all calls to DBFile.cc are now implemented in three classes, namely, GenericDBFile.cc, HeapDBFile.cc, and SortedDBFile.cc. The functionality of the Heap class has already been implemented earlier. The GenericDBFile contains virtual functions that are being implemented within the HeapDBFile and SortedDBFile classes.

The create function within the DBFile class now creates a new metadata file that contains information, which will be useful while opening the DBFile after its closed. After that it creates a new SortedDBFile or HeapDBFile class as per the inputs provided. The constructor of the SortedDBFile class takes the structure SortInfo as an argument that contains information pertaining to the runLength and the OrderMaker class.

The Open() function within the DBFile class opens the metadata file related to that fPath and creates HeapDBFile or SortedDBFile instance based the information stored in the meta data file . The myInternalVar of the GenericDBFile class then calls the functionalities implemented within the HeapDBFile or SortedDBFile class.

The functions created in the sorted class have been discussed below as the heap class is an extension of previous assignments. Please note that both the heap and sorted class have implemented the following functions:

1. SwitchToWriteMode()
2. SwitchToReadMode()
3. AddToDBFile(Record &addme)
4. GetNextFromDBFile(Record &fetchme)
5. GetNextFromDBFile(Record &fetchme, CNF &cnf, Record &literal)

SwitchToWriteMode():

It first checks if the file is currently in writing mode or reading mode. If it is in reading mode it creates a new instance of BigQ instance by passing input pipe, output pipe, sort order and run length, which will be used as differential file. After that it changes the current file mode to writing.

The SwitchToWriteMode() is called in the Add() or Load() method of the DBFile class. This maintains a differential BigQ file that takes input pipe, output pipe, sort order, and run-length as inputs. When in writing mode, the sorted DBFile inserts the new records into its instance of BigQ file and adds record to the input pipe.

If it is already in the writing mode, it just adds the record into the input pipe.

SwitchToReadMode():

The SwitchToReadMode() is called within the MoveFirst(), GetNext() and Close() functions within the GenericDBFile class. It first checks if the current file is in writing mode or reading mode. If it is in writing mode, MergeCurrentFileAndBigQOutput() private method is called. If it is already in reading mode it doesn't do anything.

Where MergeCurrentFileAndBigQOutput() method merges the current file and BigQ's output, to sort all the records.

AddToDBFile():

After switching to writing mode, we can accept new records that should be added to the BigQ's input pipe such that the newly inserted record can be sorted.

GetNextFromDBFile(Record &fetchme):

This function calls the GetRecordFromReadBufferPage() function which is present in the GenericDBFile class and is responsible for fetching the next record from the DBFile using the read page buffer.

GetNextFromDBFile(Record &fetchme, CNF &cnf, Record &literal):

This function first creates the common order maker ("query") using cnf and current file's sort order. Then it performs binary search using this query order maker to find a page such that returned page's first record is less than the literal and next page's first record is equal to or greater than to the literal. After that it sequentially it starts checking the query if query matches with the record it checks the entire cnf with the literal and if cnf returns true, it returns the current fetched record. As soon as it finds the record which is greater than the literal in terms of query, it stops and returns 0. If query order makes comes out to be empty, it does sequential search instead to find the record which satisfies cnf and literal.

Test cases:

i) Test-1: 1, 3, (c_phone), 8, 2

```
thunder:14% ./test.out
** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location:      catalog
tpch files dir:        ../tpch-dbgen/
heap files dir:        db-files/

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

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

specify sort ordering (when done press ctrl-D):
(c_phone)

specify runlength:
8

output to dbfile : db-files/customer.bin
input from file : ../tpch-dbgen/customer.tbl

select option for : db-files/customer.bin
  1. add a few (1 to 1k recs)
  2. add a lot (1k to 1e+06 recs)
  3. run some query
2
.....
added 150000 recs..so far 150000

create finished.. 150000 recs inserted
```

ii) Test-2: 2, 3

```
yogesh@yogi-pc ~
File Edit View Search Terminal Help
.c_custkey: [25985], c_name: [Customer#000025985], c_address: [n wWLOO AZTfsAYBRctoZDAoUe], c_nationkey: [24], c_phone: [34-996-827-5819], c_acctbal: [6393.05], c_mktsegment: [BUILDING], c_comment: [sly a
cross the blithely even deposits. slyly pending excuses nag slyly. packages ]
.c_custkey: [6948], c_name: [Customer#000006948], c_address: [DNtLKcdov3NhFFAGa48Pi6cY1r5] NF], c_nationkey: [24], c_phone: [34-996-984-3634], c_acctbal: [2634.46], c_mktsegment: [AUTOMOBILE], c_comment:
[ sleep blithely across the regular deposits. accounts wake. slyly ironic gifts above th]
.c_custkey: [2068], c_name: [Customer#000002068], c_address: [q3hQ8NHhSj2n], c_nationkey: [24], c_phone: [34-996-906-1652], c_acctbal: [-839.34], c_mktsegment: [HOUSEHOLD], c_comment: [ the closely expre
ss excuses wake furiously furiously special reqs]
.c_custkey: [161355], c_name: [Customer#000161355], c_address: [sJ5taX90LFQW], c_nationkey: [24], c_phone: [34-997-188-9666], c_acctbal: [32.59], c_mktsegment: [FURNITURE], c_comment: [even sheaves wake q
uickly expr]
.c_custkey: [37088], c_name: [Customer#000037088], c_address: [12kFdo3fodNGVn50MJ5s], c_nationkey: [24], c_phone: [34-997-204-5897], c_acctbal: [2484.13], c_mktsegment: [AUTOMOBILE], c_comment: [al theod
olites haggle across the quickly tr]
.c_custkey: [3174], c_name: [Customer#000003174], c_address: [9XGPFKLWZMU], c_nationkey: [24], c_phone: [34-997-244-3200], c_acctbal: [6117.29], c_mktsegment: [BUILDING], c_comment: [round the furiously e
xpress packages. furiously regular courts are furto]
.c_custkey: [86364], c_name: [Customer#000086364], c_address: [vMKeNm Ez2a75U3vknI2Rnp0jC9], c_nationkey: [24], c_phone: [34-997-251-4190], c_acctbal: [2691.67], c_mktsegment: [AUTOMOBILE], c_comment: [ e
ven foxes haggle quickly blithely regular Pinto beans. slyly regular ideas sleep slyly along the r]
.c_custkey: [35286], c_name: [Customer#000035286], c_address: [P HMQQnLb9j5J0ps], c_nationkey: [24], c_phone: [34-997-517-4660], c_acctbal: [120.81], c_mktsegment: [AUTOMOBILE], c_comment: [und the perm
anent packages sleep blithely according to the special, even requests. express foxes ]
.c_custkey: [48253], c_name: [Customer#000048253], c_address: [gAVusa09FVBERzgQoz 5lQC9wK07gYXlTLRKg], c_nationkey: [24], c_phone: [34-997-697-4316], c_acctbal: [4429.09], c_mktsegment: [AUTOMOBILE], c_c
omment: [express forges use after the carefully ironic ideas. slyly express foxes sleep quickly. carefully special]
.c_custkey: [31537], c_name: [Customer#000031537], c_address: [82Ffr5r7Gwkt2as], c_nationkey: [24], c_phone: [34-998-190-8709], c_acctbal: [-252.7], c_mktsegment: [HOUSEHOLD], c_comment: [ructions cajole l
ideas. slyly express depos]
.c_custkey: [117513], c_name: [Customer#000117513], c_address: [cuBZ TE0EONHPmo0dv6EIHxKU43AV], c_nationkey: [24], c_phone: [34-998-315-4817], c_acctbal: [2334.66], c_mktsegment: [AUTOMOBILE], c_comment:
[ c Ideas are. accounts haggle agains]
.c_custkey: [53730], c_name: [Customer#000053730], c_address: [6fv7lW0Pp9lQntCZiswle2GXoxj9A ], c_nationkey: [24], c_phone: [34-998-302-6518], c_acctbal: [4923.4], c_mktsegment: [MACHINERY], c_comment:
[ silent requests haggle carefully regular, regular packages. enticing, even deposits wake slyly along the]
.c_custkey: [12022], c_name: [Customer#000012022], c_address: [PwVDMfqZtkW0o], c_nationkey: [24], c_phone: [34-998-408-4382], c_acctbal: [2877.36], c_mktsegment: [HOUSEHOLD], c_comment: [layers. regular p
into beans about the quickly regular ideas cajole blithely]
.c_custkey: [125210], c_name: [Customer#000125210], c_address: [B0L3Ag79tVstvdXpPeb3sugKa0VV], c_nationkey: [24], c_phone: [34-998-711-2597], c_acctbal: [6678.32], c_mktsegment: [FURNITURE], c_comment: [l
y ironic, ironic accounts. theodolites among the slyly even patterns nag slyly flufft]
.c_custkey: [50667], c_name: [Customer#000050667], c_address: [9agXwCvLCGoy2eA], c_nationkey: [24], c_phone: [34-999-115-2195], c_acctbal: [8764.52], c_mktsegment: [FURNITURE], c_comment: [lly regular asy
mptotes around the slyly bold instructions boost fur]
.c_custkey: [7281], c_name: [Customer#00007281], c_address: [dZCq2e94SxQjxAd7Ww0WMyFVKGD4FIRashk], c_nationkey: [24], c_phone: [34-999-121-7718], c_acctbal: [82.07], c_mktsegment: [HOUSEHOLD], c_comme
nt: [ns sublate furiously. carefully ironic platelats grow furiously. ironic, bold theodolites hang above th]
.c_custkey: [84642], c_name: [Customer#000084642], c_address: [M3erprOnUVRK], c_nationkey: [24], c_phone: [34-999-195-7029], c_acctbal: [6912.63], c_mktsegment: [BUILDING], c_comment: [totes above the pen
ding, regular hockey players sleep blithely after the blithely even deposits. carefully regula]
.c_custkey: [148611], c_name: [Customer#000148611], c_address: [H0NKAELdZvMyGhp-e2ZgShad7F00NlH64], c_nationkey: [24], c_phone: [34-999-283-6448], c_acctbal: [24.12], c_mktsegment: [HOUSEHOLD], c_comme
nt: [requests above the blithely pending asymptotes sleep carefully about the slyly even excu]
.c_custkey: [102689], c_name: [Customer#000102689], c_address: [uPFxngv50aDQIL], c_nationkey: [24], c_phone: [34-999-363-7145], c_acctbal: [6256.97], c_mktsegment: [AUTOMOBILE], c_comment: [lly regular asy
mptotes haggle fluffly according to the furiously bold courts. daring ]
.c_custkey: [35478], c_name: [Customer#000035478], c_address: [0tSzcylLHzXXlvX4Y2FeToQih,tovs6zYA], c_nationkey: [24], c_phone: [34-999-379-6016], c_acctbal: [6097.18], c_mktsegment: [MACHINERY], c_comme
nt: [t(hout the fluffly ironic packages. daring, even de]
.c_custkey: [103505], c_name: [Customer#000103505], c_address: [XPXNpdCD7HRqCV8ABpJf3lxDc9AGX ZB2 ], c_nationkey: [24], c_phone: [34-999-404-9770], c_acctbal: [582.88], c_mktsegment: [FURNITURE], c_commen
t: [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: [DpHfJfKVKr4LQelF9t55SKzDUa4G3asKeya5N7dz], c_nationkey: [24], c_phone: [34-999-535-1203], c_acctbal: [7118.29], c_mktsegment: [FURNITURE], c_c
omment: [c requests sleep. quickly silent requests wake fluffly. furiously close packages nag furiously across the caref]
.c_custkey: [94308], c_name: [Customer#000094308], c_address: [5aE9PLelngVg8v0KCGyR,NhDTSuc4], c_nationkey: [24], c_phone: [34-999-618-6653], c_acctbal: [8636.13], c_mktsegment: [FURNITURE], c_comment:
[g, even deposits. carefully unus]
.c_custkey: [102414], c_name: [Customer#000102414], c_address: [Xzg9elWbsxIQW,LUCYUAcPoEndLUQ], c_nationkey: [24], c_phone: [34-999-618-6881], c_acctbal: [9614.62], c_mktsegment: [MACHINERY], c_comment:
[usly final packages. carefully even packages by the permanently special gifts nag]

scanned 150000 recs

thunder:31% █
```

iii) Test-3: 3,3,(c_phone > '34-999-195-7029') AND (c_mktsegment = 'FURNITURE')

```
thunder:31% ./test.out
** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location: catalog
tpch files dir: ../tpch-dggen/
heap files dir: db-files/

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
3

enter CNF predicate (when done press ctrl-D):
(c_phone > '34-999-195-7029') AND (c_mktsegment = 'FURNITURE')
.c_custkey: [103505], c_name: [Customer#000103505], c_address: [XPXNpdCD7HRqCV8ABpJf3lxDc9AGX ZB2 ], c_nationkey: [24], c_phone: [34-999-404-9770], c_acctbal: [582.88], c_mktsegment: [FURNITURE], c_c
omment: [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: [DpHfJfKVKr4LQelF9t55SKzDUa4G3asKeya5N7dz], c_nationkey: [24], c_phone: [34-999-535-1203], c_acctbal: [7118.29], c_mktsegment: [FURNITURE], c_c
omment: [c requests sleep. quickly silent requests wake fluffly. furiously close packages nag furiously across the carer]
.c_custkey: [94308], c_name: [Customer#000094308], c_address: [5aE9PLelngVg8v0KCGyR,NhDTSuc4], 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 db-files/customer.tbl returned 4 recs

thunder:32% █
```

Gtests:

Instructions to run:

1. Update catalog_path, dbfile_dir, & tpch_dir paths in “gtests/SortedDBFileTests.cc”.
2. Run “make gTestSortedDBFile.out”.
3. Run “./gTestSortedDBFile.out”.

Details:

The “TestMetaDataFile” tests the metadata file. It tests metadata file after calling the create() method of DBFile. It does the following checks.

1. It checks if the dbFile is present or not.
2. It checks if the metadata file is present or not.
3. It checks the metadata file’s content.
 - a. It checks the type stored in the first line of the file.
 - b. It checks the run length stored in the second line of the file.
 - c. It checks the sort order stored in the third line of the file.

Output:

```
dhirajmaheshparyani@Dhirajs-MacBook-Pro Project % make gTestSortedDBFile.out; ./gTestSortedDBFile.out
make: `gTestSortedDBFile.out' is up to date.
[=====] Running 1 test from 1 test case.
[-----] Global test environment set-up.
[-----] 1 test from SortedDBFileTest
[ RUN      ] SortedDBFileTest.TestMetaDataFile
[      OK   ] SortedDBFileTest.TestMetaDataFile (0 ms)
[-----] 1 test from SortedDBFileTest (0 ms total)

[-----] Global test environment tear-down
[=====] 1 test from 1 test case ran. (0 ms total)
[  PASSED  ] 1 test.
dhirajmaheshparyani@Dhirajs-MacBook-Pro Project %
```