

Report: Assignment 2 (part 2)

Implementing a Sorted File (part 2)

Group Members:

1. Smrati Pandey, UFID: 4323-9459
2. Wins Goyal, UFID: 7357-1559

Instructions for Execution / Code compilation and to run tests:

I. To extract folders

- a. Extract the contents of the folder, named “SmratiPandey_WinsGoyal_p2-2.zip”.
- b. Open the terminal.
- c. Navigate to the extracted folder.

II. To run the tests

- a) Perform following these commands, first:

```
$ make clean
$ make test.out
$ ./test.out
```

Note: Before executing the above commands, generate the *tpch* files using the *dbgen* program. If needed, also change the *tpch*, the *dbfile output* and *catalog directories*, in the *test.h* file.

- b) Post-executing above commands gives again a menu-based interface, which helps test the code through following three options. These can be selected as needed.

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

- c) Now, to perform the tests, provide appropriate inputs to the menu-based interface, table name selected, run-length and sort-order, as needed.

Summarized Documentation of the implemented methods and functions:

This part of the project demanded us to extend the *DBFile.cc* by implementing the *GenericDBFile* class and modifying & extending the *Heap.cc* and *Sorted.cc* files. Although most of the functionalities have already been summarized in the previous project reports. This report summarizes the new additions and modifications to the source code for implementation.

I. *DBFile* >> *GenericDBFile*:

- Unlike previously implemented i.e. the *DBFile* was implemented to call and create a *Heap DBFile*, now the *DBFile.h* and *DBFile.cc* are extended to use new class ‘*GenericDBFile*’ as a virtual base class that will either call *Heap* or *Sorted* file functionalities as deemed appropriate.
- *Heap* and *Sorted* are two subclasses in the *GenericDBFile* class, the instances of which are created and get stored in the internal variable called ‘*tempVal*’, defined in the *DBFile.h*, when either of following two functions are called:

A) *DBFile.Create*:

- a) The create function takes as input the filepath, filetype and startup. For this assignment, the filetype can be ‘*heap*’ or ‘*sorted*’. This class stores this type of information in a metadata file associated with the created dbfile.

- b) The created file is stored in a *GenericDBFile* type variable named 'tempVal'.
- c) If it's 'heap', the class will create an instance of the 'heap dbfile' calling the functionalities implemented in the *heap.cc*.
- d) Similarly, if it is 'sorted', the class will create the instance of 'sorted dbfile' calling the functionalities of *sorted.cc* along with passing the parameter 'startup'.
- e) The 'NewRunData' is a 'struct' type data structure that ('startup' in the requirements report) provides the *OrderMaker* and *runLength* to the Sorted and BigQ class functionalities.
 - *OrderMaker* instance stores a text version of the ordering criteria.
 - 'NewRunData' is defined in 'comparison.h'.

B) *DBFile.Open*:

- a) This class first reads from the metadata file associated with the selected *DBFile* and checks the file type.
 - b) The file to be opened is stored in *GenericDBFile* type variable named 'tempVal'.
 - c) If it's heap, this will call the functionalities of the heap class.
 - d) Otherwise if it's sorted, the class will pass on the 'OrderMaker' to the sorted class and call it's functionalities.
 - e) Because the *OrderMaker* is in text form, this class will re-instantiate the *OrderMaker* first.
- *GenericDBFile* contains the same functionalities as the *DBFile* does, just that these functionalities now are *virtual* type.

II. **Heap Class:**

- The class is called via '*GenericDBFile*' interface, and uses the following functionalities. We don't re-explain them here, as there is not much modifications to their implementation as documented in the first assignment:
 - *Add*
 - *Create*
 - *Open*
 - *Load*
 - *GetNext*
 - *MoveFirst*
 - *Close*

III. **Sorted Class** (and some modification in *BigQ* class):

- Although we have documented some of the functionalities of the Sorted class in assignment 2 (part 1) while implementing the *BigQ* class, here we summarize the modified and extended implementation of this class.
- The class has an enum variable of type *Dirty* i.e. '*isDirtyVal*' store two modes for the created/opened dbfile: I. '*Read*' ; II. '*Write*'.
- In writing mode, a *BigQ* instance is also generated (provided with the *runLength* when *DBFile.Create* called it) along with the sorted *DBFile* wrapped inside the File class instance.
- In reading mode *BigQ* is not instantiated.
- Depending on this mode, the different functions are called for a particular *DBFile* or same function would work differently for different modes.
 - *Load*: The function loads the *DBFile*, reads its records according to the *OrderMaker* given to *BigQ* class, when in writing/reading mode.
 - *Add*: The function adds new record(s) into the *BigQ* instance from the records reads from *DBFile*, when in writing mode. In reading mode, an internal *BigQ* instance will be there within the *DBFile* to add the new records to *BigQ* before quickly changing the mode 'writing'.

- *Create*: It will create the *DBFile* at the filepath, if not already there with the ‘*Read*’ mode.
- *Open*: It will open the *DBFile* at the filepath, with the ‘*Read*’ mode.
- *Merge*: This function will be called before *MoveFirst*, *GetNext* or *Close* functions. This function empties the internal *BigQ* instance by merging it with the sorted data before changing to ‘*Read*’ mode.
 - i. Input pipe to *BigQ* is blocked and the records are taken from the Output pipe as per *OrderMaker* criteria
 - ii. In parallel, sorted file is also read as per the *OrderMaker* criteria.
 - iii. The two data are merged using the 2-way merge algorithm as implemented in assignment 2-1, to a new File instance. This file is now available for ‘*Read*’ to be queried or for ‘*Close*’.
- Following functions accept *CNF* instance and are used for ‘Query processing’.
 - *MoveFirst*: If the mode is ‘*write*’, first calls the *Merge*, then converts to ‘*Read*’ mode. This function puts the pointer at the first record of the file.
 - *GetNext (with/without CNF)*: If the mode is ‘*write*’, first calls the *Merge*, then converts to ‘*Read*’ mode. This function returns the next record in the file through ‘*Binary Search*’ implementation that satisfies the *CNF* condition (‘*selection predicate*’) after checking where the sorting order matches.
 - i. ‘*query OrderMaker*’ is created from the *CNF*. In implementation, this is ‘*givenSortedOrder*’, which is implemented in ‘*comparison.cc*’ and being used by *BigQ* and *Sorted file* calls.
 - ii. This uses ‘*newSortedOrderGiven*’ defined in *comparison.h*
 - iii. The records are read and compared after the current Record in the *givenSortedOrder*.
 - iv. ‘*query OrderMaker*’ is created by matching the ordering attributes of the records with the *CNF* disjunctions in its instance. Attributes literals are added sequentially to the query until a match fails. This is used as the comparison criteria in the *Binary Search*.
 - v. After *Binary Search*, first ‘*query OrderMaker*’, then ‘*CNF*’ are evaluated w.r.t. to the records obtained. If *queryOrderMaker* accepts the record but the *CNF* doesn’t, the next record in line then evaluated for both.
 - vi. ‘*queryOrderMaker*’ is same for repeated calls to *GetNext*.
 - *fetchVal*: This function fetches values as per the *CNF* conditions. Object ‘*compareVal*’ of type ‘*ComparisonEngine*’ is used for comparing.
 - *newFetchVal*: This works almost same as the *fetchVal*, just that it also uses ‘*givenSortedOrder*’ for the comparison.
 - *Close*: This closes the *DBFile* after writing the metadata file to it and ensuring that the internal *BigQ* was appropriately merged with the file.
- Changes in “*comparison.cc*”:

OrderMaker is modified and extended for the *givenSortedOrder* to generate query *OrderMaker* and for generating the meta data files for *DBFiles*. One definition is about getting all the ordering attributes while other definition is to create a meta file in which those ordering attributes will be written.
- Changes in “*file.h*” & “*file.cc*”:
 - Class ‘*LengthGen*’ defined in *File.h* with public method ‘*GetNext*’. This generates the ‘*runLength*’ for a file. *LengthGen* takes as input the *start* and *end* for the *DBFile* as well as the *DBFile*.
 - In *File.cc*, this class implements the public method ‘*GetNext*’ that takes a pointer of class *Record* to the first record in that *DBFile*. The method has the procedure to find the next record after the current.

- Class 'NewGeneration' is also defined in *File.h* that has public class variable 'run' of type *LengthGen*. This takes as input *number of records* and the *runLength* in its constructor.
- *Page* class implements the 'getValue' method in *File.cc* which is '*currentSizeInBytes – sizeof(int)*'.

Screenshots of Running All Test Cases

Perform following commands to run all the four test cases:

```
$ make clean
$ make test.out
$ ./test.out
```

Test case 1:

1
3
(c_phone)
8

```
** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location:      catalog
tpch files dir:        /home/smrati/Downloads/a2-2test/tpch-dbgen/
heap files dir:         /home/smrati/Downloads/a2-2test/bin/

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 : /home/smrati/Downloads/a2-2test/bin/customer.bin
input from file : /home/smrati/Downloads/a2-2test/tpch-dbgen/customer.tbl

select option for : /home/smrati/Downloads/a2-2test/bin/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
```

Test case 2:

2
3

```
Terminal: Local x +
.c_custkey: [6048], c_name: [Customer#00006048], c_address: [DNtLkGdov3MnFfAGa40Pi6c6Yr5j NF], c_nationkey: [24], c_phone: [34-996-904-3634], c_acctbal: [2634.46], c_mktsegment: [AUTOMOBILE], c_comment: [ sleep blithely across the regular deposits. a
unts wake. slyly ironic gifts above th]
.c_custkey: [2068], c_name: [Customer#00002068], c_address: [q3hQ8nHhnsJ2n], c_nationkey: [24], c_phone: [34-996-906-1652], c_acctbal: [-839.34], c_mktsegment: [HOUSEHOLD], c_comment: [ the closely express excuses wake furiously furiously special requ
.c_custkey: [101355], c_name: [Customer#000101355], c_address: [sJ5taX90[FOW], c_nationkey: [24], c_phone: [34-997-188-9666], c_acctbal: [32.59], c_mktsegment: [FURNITURE], c_comment: [even sheaves wake quickly expr]
.c_custkey: [37088], c_name: [Customer#000037088], c_address: [I2kfDo3fodNGYonS0MJs5], c_nationkey: [24], c_phone: [34-997-204-5897], c_acctbal: [2484.13], c_mktsegment: [AUTOMOBILE], c_comment: [al theodolites haggle across the quickly ir]
.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 express packages. furiously regular courts are fi
]
.c_custkey: [86364], c_name: [Customer#000086364], c_address: [vMKmNm EzZa75U3vkmI2Rnp0jC9], c_nationkey: [24], c_phone: [34-997-251-4190], c_acctbal: [2691.67], c_mktsegment: [AUTOMOBILE], c_comment: [ even foxes haggle quickly blithely regular pinto
ns. slyly regular ideas sleep slyly along the fl]
.c_custkey: [35286], c_name: [Customer#000035286], c_address: [P HN6QXTmI8pJ5Xps], c_nationkey: [24], c_phone: [34-997-517-4660], c_acctbal: [120.81], c_mktsegment: [AUTOMOBILE], c_comment: [und the permanent packages sleep blithely according to the s
ial, even requests. express foxes ]
.c_custkey: [48253], c_name: [Customer#000048253], c_address: [gAvUsA09FVBERzq0oz 51Qc9wYk07gYX1TLRkG], c_nationkey: [24], c_phone: [34-997-697-4316], c_acctbal: [4429.09], c_mktsegment: [AUTOMOBILE], c_comment: [express forges use after the carefully
nic ideas. slyly express foxes sleep quickly. carefully special]
.c_custkey: [31537], c_name: [Customer#000031537], c_address: [82Fr5r7gWxt2a5], c_nationkey: [24], c_phone: [34-998-190-8709], c_acctbal: [-252.7], c_mktsegment: [HOUSEHOLD], c_comment: [ructions cajole ideas. slyly express depos]
.c_custkey: [117513], c_name: [Customer#000117513], c_address: [cuBZ TE8EONnPmoOdv6EIhXU43AV], 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: [6fv7IW9Pr91QntC2zwsile2GX0xj9A ], c_nationkey: [24], c_phone: [34-998-382-6518], c_acctbal: [4923.4], c_mktsegment: [MACHINERY], c_comment: [ silent requests haggle carefully regular, regu
packages. enticing, even deposits wake slyly among the]
.c_custkey: [12022], c_name: [Customer#00012022], c_address: [PwDMfqZtkW0o], c_nationkey: [24], c_phone: [34-998-408-4302], c_acctbal: [2077.36], c_mktsegment: [HOUSEHOLD], c_comment: [layers. regular pinto beans about the quickly regular ideas cajol
lithely]
.c_custkey: [125210], c_name: [Customer#000125210], c_address: [B3iJAg79TVtsDxgPeb63sUqKa0Vv], c_nationkey: [24], c_phone: [34-998-711-2597], c_acctbal: [6678.32], c_mktsegment: [FURNITURE], c_comment: [ly ironic, ironic accounts. theodolites among the
lyly even patterns nag slyly fluffi]
.c_custkey: [50667], c_name: [Customer#000050667], c_address: [9aGwxCYLcG6y2eA], c_nationkey: [24], c_phone: [34-999-115-2195], c_acctbal: [8764.52], c_mktsegment: [FURNITURE], c_comment: [mptotes around the slyly bold instructions boost furil]
.c_custkey: [7288], c_name: [Customer#000007288], c_address: [,dZCqJ94SsqXQjxAd7WxRMyyFVKgD4pFTrAshk], c_nationkey: [24], c_phone: [34-999-121-7718], c_acctbal: [82.07], c_mktsegment: [HOUSEHOLD], c_comment: [ns sublate furiously. carefully ironic plat
ts grow furiously. ironic, bold theodolites hang above th]
.c_custkey: [84642], c_name: [Customer#000084642], c_address: [MJerprOnUYRK], c_nationkey: [24], c_phone: [34-999-195-7029], c_acctbal: [6912.63], c_mktsegment: [BUILDING], c_comment: [totes above the pending, regular hockey players sleep blithely afte
he blithely even deposits. carefully regula]
.c_custkey: [148611], c_name: [Customer#000148611], c_address: [HoNKAEld2vWMyGmpw3ZZgSNdS7f0dNIH64], c_nationkey: [24], c_phone: [34-999-283-6448], c_acctbal: [24.12], c_mktsegment: [HOUSEHOLD], c_comment: [requests above the blithely pending asympt
sleep carefully about the slyly even excu]
.c_custkey: [102689], c_name: [Customer#000102689], c_address: [uPFxngvS0aDQIL], c_nationkey: [24], c_phone: [34-999-363-7145], c_acctbal: [6256.97], c_mktsegment: [AUTOMOBILE], c_comment: [ly regular asymptotes haggle fluffily according to the furious:
bold courts. daring ]
.c_custkey: [35478], c_name: [Customer#000035478], c_address: [Ot5czyiiLhzXXivx4Y2FeToQIH,i0vs6zyA], c_nationkey: [24], c_phone: [34-999-379-6016], c_acctbal: [6097.18], c_mktsegment: [MACHINERY], c_comment: [lithout the fluffily ironic packages. daring
ven de]
.c_custkey: [103505], c_name: [Customer#000103505], c_address: [XPNXpdCD7HrqcV8ABpjF3ixDc9AGX 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 re
sts could have to are according to the final accounts. final]
.c_custkey: [62707], c_name: [Customer#000062707], c_address: [bRbtSTJXhh], 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: [DpHfJFKVkr4LQeif9t5SSKzDUa4G3asKeya5M7dz], c_nationkey: [24], c_phone: [34-999-535-1203], c_acctbal: [7118.29], c_mktsegment: [FURNITURE], c_comment: [c requests sleep. quickly silent requ
wake fluffily. furiously close packages nag furiously across the caref]
.c_custkey: [94308], c_name: [Customer#000094308], c_address: [5aE9PLeNgeIvg0v6KCg0yR,NhDT5uC4], 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: [Xzg9eiMbsxiQW,LUCyUAcPooENDLUQt], c_nationkey: [24], c_phone: [34-999-618-6881], c_acctbal: [9614.62], c_mktsegment: [MACHINERY], c_comment: [usly final packages. carefully even packages t
he permanently special gifts ne]

scanned 150000 recs
```

Test case 3:

3
3

(c_phone > '34-999-195-7029') AND (c_mktsegment = 'FURNITURE')

```
smrati@smrati-XPS-15-7590:~/Downloads/a2-2test$ ./test.out
```

**** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT ****

```
catalog location:    catalog
tpch files dir:      /home/smrati/Downloads/a2-2test/tpch-dbggen/
heap files dir:       /home/smrati/Downloads/a2-2test/bin/
```

```
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: [XPNXpdCD7HRqcV8ABpjF3ixDc9A6X ZB2 ], c_nationkey: [24], c_phone: [34-999-404-9770], c_acctbal: [582.88], c_mktsegment: [FURNITURE], c_comment: [fully even deposits boost. quickly iror
c requests could have to are according to the final accounts. final]
c_custkey: [62707], c_name: [Customer#000062707], c_address: [bRbtSTJXhh], 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: [DpHfJFKVkr4LQeif9t5SSKzDUa4G3asKeya5M7dz], c_nationkey: [24], c_phone: [34-999-535-1203], c_acctbal: [7118.29], c_mktsegment: [FURNITURE], c_comment: [c requests sleep. quickly silent request]
wake fluffily. furiously close packages nag furiously across the caref]
c_custkey: [94308], c_name: [Customer#000094308], c_address: [5aE9PLeLngeVg8v6KCg0yR,Nh0TSu4], 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 /home/smrati/Downloads/a2-2test/bin/customer.bin returned 4 recs

GTests:

Perform following commands to executing the GTests:

```
$ make clean
$ make myGtest.out
$ ./myGtest.out
```

- OpenSortHeap*: This calls the 'Open' function in Sorted file, to test whether the file has been successfully opened or not.
- CreateSortHeap*: This calls the 'create' function in Sorted file, to test whether the file has been successfully created or not.

Following is the screenshots of all these two test cases in execution (terminal):

```
Terminal: Local x +
smrati@smrati-XPS-15-7590:~/Downloads/a2-2test$ ./myGtest.out
[=====] Running 2 tests from 2 test cases.
[-----] Global test environment set-up.
[-----] 1 test from OpenSortHeap
[ RUN      ] OpenSortHeap.OpenSortHeap_Test
[ OK       ] OpenSortHeap.OpenSortHeap_Test (0 ms)
[-----] 1 test from OpenSortHeap (1 ms total)

[-----] 1 test from CreateSortHeap
[ RUN      ] CreateSortHeap.SuccessCreateFile_Test
[ OK       ] CreateSortHeap.SuccessCreateFile_Test (0 ms)
[-----] 1 test from CreateSortHeap (0 ms total)

[-----] Global test environment tear-down
[=====] 2 tests from 2 test cases ran. (1 ms total)
[ PASSED  ] 2 tests.
smrati@smrati-XPS-15-7590:~/Downloads/a2-2test$
```