Project Report 2

Part 2

Implementing a Sorted File

Arpan Banerjee

UFID: 9359-9083

arpanbanerjee@ufl.edu

Krutantak Patil

UFID: 5615-6343

Krutantakb.patil@ufl.edu

1) Steps to compile and run - code, tests and gtests

*(*Assuming .bin files would be provided to us in <u>dbfiles</u> folder)

  i.   **make main** – Command to build the main program.
  ii. **./main** – Command to run the main executable.
  iii. **make test.out** – Command to build the tests.
  iv. **./test.out** – Command to run the tests.
  v.   **make dBFileTest.out** – Command to build the gtests.
  vi. **./gtestScript.sh** – Command to run the gtests. (<u>gtest</u> folder contains data required for running all 12 test cases written in DBFileTest.cc file)

2) Sorted file implementation

Data members

  i.   **int runlen** – Used to store length of each run.
  ii. **File runFile** – Used to store records and pages in the file.
  iii. **OrderMaker sortOrder** – Pointer to page holding recent record additions to the file.
  iv. **Page p** – Used to temporarily store records before adding into the input pipe and also while getting records using GetNext.
  v.   **mode fileMode** – Stores the current mode of the file as an enum. Has values either read or write.

Member functions

  i.   **ToggleMode** (mode newMode)

       a. *Used to switch the mode of the file to read or write.*
       b. *If switching to read mode: Two way merging of the BigQ (diff file) and existing sorted file occurs.*
       c. *If switching to write mode: We clear the page object to enable writing into it.*

  ii. **producerWorker** (void* arg)

       a. Worker routine used to load data into input pipe for BigQ -
       b. Used in the Load() function to read all records from given file and insert into input pipe.

c. Also used for merging, wherein all the records from the temporary page are inserted into the pipe.

iii. **consumerWorker** (void *arg)

a. It is a worker routine which reads the data from the ouput pipe and inserts into pages in a file.

iv. **TwoPassMergeing** ()

a. This function first creates threads for the producer and consumer workers.
b. It then initializes a BigQ instance which sorts in runs and merges them, releasing records in the output pipe.
c. The threads are then joined completing the merge process.

v. **Create** (char *filePath, fileTypeEnum fileType, void *startup)

a. This creates a new sorted file using the runLength and OrderMaker provided in the startup argument.
b. It then sets the file mode to write.

vi. **Load** (Schema &fileSchema, char *loadPath)

a. This function loads data from the given loadPath argument and uses the worker threads as described before.

vii. **Open** (char *filePath)

a. This function opens an existing sorted file.
b. The file mode is set to read by default.
c. The seek is moved to the start using MoveFirst.

viii. **MoveFirst** ()

a. This first toggles the file mode to read and then sets the read page as the first page.

ix. **Add** (Record &rec)

a. This first switches the mode to write and then adds the record to the temporary page.

x. **GetNext** (Record &fetchMe)

a. This function first toggles the mode to read.

    b. Then the page is loaded using GetPage() and records are extracted using GetFirst().

   xi. **GetNext** (Record &fetchMe, CNF &cnf, Record &literals)

    a. This is similar to GetNext without a CNF. The records are compared against the CNF along with the optimizations mentioned in the document.

3) ./test.out results for 1GB data

  i. 1
3
(c_phone)
8
2

```
arpan@arpan-laptop:~/Desktop/Krutantak/Git/database-system-implementation/a2$ ./test.out

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

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 : ../dbfiles/customer.bin
input from file : ../tpch-dbgen/customer.tbl

select option for : ../dbfiles/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
arpan@arpan-laptop:~/Desktop/Krutantak/Git/database-system-implementation/a2$
```

ii.　2
　　　3

```
E], c_comment: [ sleep blithely across the regular deposits. accounts wake. slyly 1
.c_custkey: [2068], c_name: [Customer#000002068], c_address: [q3hQ8nHhhsJ2n], c_nat
 closely express excuses wake furiously furiously special requ]
.c_custkey: [101355], c_name: [Customer#000101355], c_address: [sJStaX90lFOW], c_na
sheaves wake quickly expr]
.c_custkey: [37088], c_name: [Customer#000037088], c_address: [I2kfDo3fodNGYonS0MJs
ent: [al theodolites haggle across the quickly ir]
.c_custkey: [3174], c_name: [Customer#000003174], c_address: [9XGPFKLWZMU], c_natio
he furiously express packages. furiously regular courts are furio]
.c_custkey: [86364], c_name: [Customer#000086364], c_address: [vMKeNm EzZa75U3vkmI2
c_comment: [ even foxes haggle quickly blithely regular pinto beans. slyly regular
.c_custkey: [35286], c_name: [Customer#000035286], c_address: [P  HN6QXTmlBpJ5Xps],
 [und the permanent packages sleep blithely according to the special, even requests
.c_custkey: [48253], c_name: [Customer#000048253], c_address: [gAVusA09FVBERzgQoz S
TOMOBILE], c_comment: [express forges use after the carefully ironic ideas. slyly e
.c_custkey: [31537], c_name: [Customer#000031537], c_address: [82Fr5r7gWxt2a5], c_n
tions cajole ideas. slyly express depos]
.c_custkey: [117513], c_name: [Customer#000117513], c_address: [cuBZ TE0EONnPFmoOdv
E], c_comment: [c ideas are. accounts haggle agains]
.c_custkey: [53730], c_name: [Customer#000053730], c_address: [6fv7IW9Pr91QntCZzwsi
], c_comment: [ silent requests haggle carefully regular, regular packages. enticin
.c_custkey: [12022], c_name: [Customer#000012022], c_address: [PwVDMfqZtKWOo], c_na
ers. regular pinto beans about the quickly regular ideas cajole blithely]
.c_custkey: [125210], c_name: [Customer#000125210], c_address: [BJiJAg79TvSTdxgPeb6
 c_comment: [ly ironic, ironic accounts. theodolites among the slyly even patterns
.c_custkey: [50667], c_name: [Customer#000050667], c_address: [9aGxwCYLcG6y2eA], c_
ptotes around the slyly bold instructions boost furi]
.c_custkey: [7288], c_name: [Customer#000007288], c_address: [,dZCqJe94SqXQjxAd7WxR
HOLD], c_comment: [ns sublate furiously. carefully ironic platelets grow furiously.
.c_custkey: [84642], c_name: [Customer#000084642], c_address: [MJerprOnUYRK], c_nat
 above the pending, regular hockey players sleep blithely after the blithely even o
.c_custkey: [148611], c_name: [Customer#000148611], c_address: [HoNNKAELd2vMNyGWpw3
HOLD], c_comment: [requests above the blithely pending asymptotes sleep carefully a
.c_custkey: [102689], c_name: [Customer#000102689], c_address: [uPFxngvSOaDQIL], c
ly regular asymptotes haggle fluffily according to the furiously bold courts. darin
.c_custkey: [35478], c_name: [Customer#000035478], c_address: [Ot5zcyiiLhzXXivX4Y2F
NERY], c_comment: [ithout the fluffily ironic packages. daring, even de]
.c_custkey: [103505], c_name: [Customer#000103505], c_address: [XPNXpdCD7HRqcV8ABpj
URE], c_comment: [fully even deposits boost. quickly ironic requests could have to
.c_custkey: [62707], c_name: [Customer#000062707], c_address: [bRbtSTIXhh], c_natio
lyly after the slyly ironic pinto beans. bold, fina]
.c_custkey: [41774], c_name: [Customer#000041774], c_address: [DpHfJFKVkr4LQeif9t5S
FURNITURE], c_comment: [c requests sleep. quickly silent requests wake fluffily. fu
.c_custkey: [94308], c_name: [Customer#000094308], c_address: [5aE9PLelngeVg8v6KCgO
], c_comment: [g, even deposits. carefully unus]
.c_custkey: [102414], c_name: [Customer#000102414], c_address: [Xzg9eiWbsxiQW,LUCyU
Y], c_comment: [usly final packages. carefully even packages by the permanently spe

 scanned 150000 recs

arpan@arpan-laptop:~/Desktop/Krutantak/Git/database-system-implementation/a2$
```

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

```
arpan@arpan-laptop:~/Desktop/Krutantak/Git/database-system-implementation/a2$ ./test.out
** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
 catalog location:      catalog
 tpch files dir:    ../tpch-dbgen/
 heap files dir:    ../dbfiles/

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 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 s
lyly 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 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 ../dbfiles/customer.bin returned 4 recs

arpan@arpan-laptop:~/Desktop/Krutantak/Git/database-system-implementation/a2$
```

5) gTest results:

i.   **TEST(DBFileTest, Sorted_Create_Equal) –**
   a. A sorted Dbfile is created using nation.tbl. This checks the
   Create function.

ii.   **TEST(DBFileTest, Sorted_Create_NotEqual) –**
   a. This function expects an exit when no table filename is
   provided to Create.

iii.   **TEST(DBFileTest, Sorted_Open_Equal) –**
   a. This function tests the Open function by opening region.bin
   (must be created beforehand).

iv.   **TEST(DBFileTest, Sorted_Close_Empty_File) –**
   a. This function tests the Close function by creating a new
   sorted DBFile and then closing it.

v.   **TEST(DBFileTest, Sorted_Close_Nations_DBFile) –**
   a. This function creates a sorted DBFile, loads data from
   nation.tbl and then closes it.

```
arpan@arpan-laptop:~/Desktop/Krutantak/Git/database-system-implementation/a2$ ./gtestScript.sh

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


 specify sort ordering (when done press ctrl-D):
        [==========] Running 10 tests from 1 test case.
[----------] Global test environment set-up.
[----------] 10 tests from DBFileTest
[ RUN      ] DBFileTest.Heap_Create_Equal
[       OK ] DBFileTest.Heap_Create_Equal (0 ms)
[ RUN      ] DBFileTest.Heap_Create_NotEqual
[       OK ] DBFileTest.Heap_Create_NotEqual (2 ms)
[ RUN      ] DBFileTest.Heap_Open_Equal
[       OK ] DBFileTest.Heap_Open_Equal (1 ms)
[ RUN      ] DBFileTest.Heap_Close_Empty_File
[       OK ] DBFileTest.Heap_Close_Empty_File (10 ms)
[ RUN      ] DBFileTest.Heap_Close_Nations_DBFile
[       OK ] DBFileTest.Heap_Close_Nations_DBFile (0 ms)
[ RUN      ] DBFileTest.Sorted_Create_Equal
[       OK ] DBFileTest.Sorted_Create_Equal (1 ms)
[ RUN      ] DBFileTest.Sorted_Create_NotEqual
[       OK ] DBFileTest.Sorted_Create_NotEqual (2 ms)
[ RUN      ] DBFileTest.Sorted_Open_Equal
[       OK ] DBFileTest.Sorted_Open_Equal (0 ms)
[ RUN      ] DBFileTest.Sorted_Close_Empty_File
[       OK ] DBFileTest.Sorted_Close_Empty_File (5 ms)
[ RUN      ] DBFileTest.Sorted_Close_Nations_DBFile
[       OK ] DBFileTest.Sorted_Close_Nations_DBFile (5 ms)
[----------] 10 tests from DBFileTest (27 ms total)

[----------] Global test environment tear-down
[==========] 10 tests from 1 test case ran. (27 ms total)
[  PASSED  ] 10 tests.
arpan@arpan-laptop:~/Desktop/Krutantak/Git/database-system-implementation/a2$
```