

# DB Implementation Project 2: Part 2

## Contributors :

*Megha Nagarmunoli UFID: 6768-3882*

*Siddharth Jain UFID: 7185-1161*

## Steps to execute:

1. Unzip the file SiddharthJain\_MeghaNagarmunoli\_p2\_2.zip

2. cd SiddharthJain\_MeghaNagarmunoli\_p2\_2/

Note: SiddharthJain\_MeghaNagarmunoli\_p2\_2 and **bin** and **tpch-dbggen** should be in the same folder/level. bin

## To run the gTests,

1. make gtest2.out

2. ./gtest2.out

Results to test test cases using the .tbl files generated from tpch-dbggen with the option -s 1 (1GB files)

Test 1:

i) 1

3

(c\_phone)

8

2

```

meghan@meghan-Inspiron-5584: ~/Documents/gitWorkspace/database-system-implementation-2/a2test
File Edit View Search Terminal Help
meghan@meghan-Inspiron-5584:~/Documents/gitWorkspace/database-system-implementation-2/a2test$ ./test2.out

** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location:      catalog
tpch files dir:        ../../tpch-dbgen/
heap files dir:        ../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 : ../bin/customer.bin
../bin/customer.bin.meta
input from file : ../../tpch-dbgen/customer.tbl

select option for : ../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

DB Open
File mode:s
.....Proc149977
Number of records1398
Setting Indexes
Begin Merge
Total 149977
File moved successfully

added 149977 recs...so far 149977

create finished.. 149977 recs inserted

meghan@meghan-Inspiron-5584:~/Documents/gitWorkspace/database-system-implementation-2/a2test$

```

Test 2:

ii) 2

3

```
meghan@meghan-Inspiron-5584:~/Documents/gitWorkspace/database-system-implementation-2/a2test$ ./test2.out
```

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

```
catalog location:    catalog
tpch files dir:      ../../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

```
.c_custkey: [117513], c_name: [Customer#000117513], c_address: [cUBZ TE6EUNPFI000V6IHXK043AV], 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: [6fv7IW9Pr91QntCZzws1e2GX0xJj9A ], c_nationkey: [24], c_phone: [34-998-382-6518], c_acctbal: [4923.4], c_mktsegment: [MACHINERY], c_comment: [ silent requests haggle carefully regular, regular packages. enticing, even deposits wake slyly among 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: [BJlJAg79TVsTdxgPeb63sUqKaOVv], 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 fluffl]
.c_custkey: [50667], c_name: [Customer#000050667], c_address: [9aGXWCYLcG6y2eA], c_nationkey: [24], c_phone: [34-999-115-2195], c_acctbal: [8764.52], c_mktsegment: [FURNITURE], c_comment: [nptotes around the slyly bold instructions boost furl]
.c_custkey: [7288], c_name: [Customer#000007288], c_address: [.dZCqJe945sqXQjxAd7WxRMVvFVKGD4pFIrAshk], c_nationkey: [24], c_phone: [34-999-121-7718], c_acctbal: [82.07], c_mktsegment: [HOUSEHOLD], c_comme nt: [ns sublate furiously. carefully ironic platelets 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 pen ding, regular hockey players sleep blithely after the blithely even deposits. carefully regula]
.c_custkey: [148611], c_name: [Customer#000148611], c_address: [HoNNKAELd2vMNYGHPw3ZZgSNaDs7fOdNIH64], 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: [uPFxngvSOaDQIL], c_nationkey: [24], c_phone: [34-999-363-7145], c_acctbal: [6256.97], c_mktsegment: [AUTOMOBILE], c_comment: [ly regular asy mptotes haggle fluffily according to the furiously bold courts. daring ]
.c_custkey: [35478], c_name: [Customer#000035478], c_address: [Ot5zcytlLhzXXlvX4Y2FeToQIh,lovS6zyA], c_nationkey: [24], c_phone: [34-999-379-6016], c_acctbal: [6097.18], c_mktsegment: [MACHINERY], c_comme nt: [lthout the fluffily ironic packages. daring, even de]
.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_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: [DpHfJFKVkr4LQelf9t5SSKzDUa4G3asKeya5M7dz], 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: [5aE9PLelngeVg8vKCG0yR,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: [Xzg9eLWbsxiQM,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 by the permanently special gifts ne]

scanned 149977 recs
```

Test 3:

iii) 3

3

(c\_phone > '34-999-195-7029') AND (c\_mktsegment = 'FURNITURE')

```
meghan@meghan-Inspiron-5584:~/Documents/gitWorkspace/database-system-implementation-2/a2test$ ./test2.out
```

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

```
catalog location:    catalog
tpch files dir:      .././tpch-dbggen/
heap files dir:      ../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')
```

```
test3
DB Open
File mode:s
  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]
Count1
  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]
Count2
  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]
Count3
  c_custkey: [94308], c_name: [Customer#000094308], c_address: [SaE9PLelNgeVg8v6KCG0yR,NhDT5uC4], c_nationkey: [24], c_phone: [34-999-618-6653], c_acctbal: [8636.13], c_mktsegment: [FURNITURE], c_comment: [g, even deposits. carefully unus]
Count4

query over ../bin/customer.bin returned 4 recs
```

```
meghan@meghan-Inspiron-5584:~/Documents/gitWorkspace/database-system-implementation-2/a2test$
```

## Result to gTest:

Testing that new records are added to Sorted DB File correctly in the sorted manner. Using a sorted bin file with 10 records and adding 10 more records to it. Then testing that all the 20 records are sorted.

```
meghan@meghan-Inspiron-5584:~/Documents/gitWorkspace/database-system-implementation-2/a2test$ ./gtest2.out
```

```
[=====] Running 1 test from 1 test case.
[-----] Global test environment set-up.
[-----] 1 test from testing_merge_code
[ RUN      ] testing_merge_code.testingMerge
.././tpch-dbggen/
input from file : ./testdata/nation.tbl
./testdata/nation.bin
DB Open
File mode:s
Number of records10
Setting Indexes
Begin Merge
Total 20
File moved successfully

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

DB Open
File mode:s
Read 20 recs from the newly merged file
[ OK ] testing_merge_code.testingMerge (2 ms)
[-----] 1 test from testing_merge_code (2 ms total)

[-----] Global test environment tear-down
[=====] 1 test from 1 test case ran. (2 ms total)
[ PASSED ] 1 test.
meghan@meghan-Inspiron-5584:~/Documents/gitWorkspace/database-system-implementation-2/a2test$
```



## Code Implementation:

### For Sorted DB File:

#### FileModes

This time we have introduced the mode in which the file is being operated i.e **readMode** or **writeMode**. If the file is in **readMode** then the functions `Open()`, `MoveFirst()`, `Close()`, `GetNext()` behave as they are supposed to behave as they would in `HeapDBFile`.

If the mode is **writeMode** and if the methods like `Add()` or `Load()` are called, then the newly added data is not appended directly to the file, it is added to an input pipe, which sorts the added data using `BigQ` (implemented in previous assignment) and the sorted data can be extracted out of the `outputPipe`.

If the data is added (`Add()` or `Load()` called) in **readMode**, then the mode is switched to **writeMode** then the newly added data is not appended directly to the file, first `BigQ` is initialized and the data added to an input pipe and the sorted data can be extracted out of the `outputPipe`.

If the mode is **writeMode** and if the methods like `Add()` or `Load()` are called, data is simply added to `inputPipe`.

If the mode is required to be changed from **writeMode** to **readMode** ( when you call functions like `Open()`, `MoveFirst()`, `Close()`, `GetNext()`) the following steps are taken to achieve this

1. Input pipe's `shutdown()` method is called.
2. Then we start to read the **.bin** file from the start ( this file is already sorted) and extract the records from the output pipe. We then merge the records from two sorted sources and write the data to an intermediate file named **mergeFile.bin**
3. Then we delete the old file and rename the **mergeFile.bin** file to the old file's name.
4. We then reset the file descriptor pointers.
5. Switch mode to **readMode**

After switching the mode, the normal operations of these methods are performed.

This is just the basic workflow of how the modes affect the functioning of the `DBFile`.

```
class GenericDBFile
```

This is a virtual class which is extended by `HeapDBFile` and `SortedDBFile`. The `create` method takes in the **f\_type** and decides which type of `DBFile` to create and is

casted to GenericDBFile. Hence, when we call any other method like **Close** or **MoveFirst**, the method of the corresponding class `HeapDBFile` or `SortedDBFile` is called.

1] `int DBFile::Create(const char *f_path, fType f_type, void *startup)`

This function determines whether to create a Heap DB or a Sorted DB File using `f_type`. It stores the type of the file at `<schema>.meta`. It calls the `Create` for `HeapDBFile` or `SortedDBFile`.

The `HeapDBFile::Create` is unchanged.

It doesn't touch information related to the **\*startup**. It is passed as such to the create method of **HeapDBFile** or **SortedDBFile**.

```
int SortedDBFile::Create (const char *f_path, fType f_type, void
*startup)
```

This function saves the `sortOrder` and `runlength` obtained from `startup` into `<schema>.metaSorted` and creates the file.

2] `int SortedDBFile::Open (const char *f_path)`

This function opens up the `<schema>.metaSorted` file and extract the `startup` information from it. It initializes the below struct with it

```
struct SortInfo { OrderMaker *myOrder; int runLength; };
```

3] `int SortedDBFile::GetNext (Record &fetchme, CNF &cnf, Record &literal) {`

This function first merges the data in `Pipe` to sorted `File` if the mode is `Write`.

It maintains a boolean `getNextCalledBefore`. If true, the function was called before and we compare the next record with `CNF` and return.

If false, scan the first page and return the record if found. If not found, we run `Binary Search` on the remaining pages by comparing first records of the pages.

4] `int GetNumofRecordPages();`

This method gets the total number of pages of the currently opened file.

5] `void AddRecordToDiskFile(File &tempFile, Page &tempPage, Record &rec, int &tempPageCount);`

This method adds the record to the page. If the page is full then it is written to the disk and the new record is added to it. This also increment the **tempPageCount** if the new page was full and written onto the disk.

6] `void get_next_Record();`

This method uses the internal file and page objects to get the next record. If the page is empty then the new page is read from the file and the first record is returned on the internal record variables.