# **Project Report: Database System Implementation (COP6726)**

**Project 2 Part 2: Extending DBFile Class** 

Name: Vikas Chaubey, UFID: 3511 5826, Email: vikas.chaubey@ufl.edu

I have done this project alone without any project partner. Hence, I am submitting it on my behalf.

**1) Compile and Run Instructions:** This program has been compiled, run and tested on Ubuntu 18.04 operating system.

#### Step.1 - Run test.out:

In order to run **test.out** first we need to make sure that in test.h file dbfile\_dir, tpch\_dir and catalog\_path is also updated correctly with right locations. please use below commands sequentially and follow onscreen instruction to run **test.out** file.

- > make clean
- > make test.out
- > ./test.out

## Step.2 - To run gtest (google tests):

In order to run **gtest** first we need to make sure that in "gtest.cc" file dbfile\_dir, tpch\_dir and catalog\_path is also updated correctly with right locations. Pass gtest\_arg.txt as argument to run gtests. This file I present in the folder. please use below commands sequentially to run gtest.

- > make clean
- > make gtest.out
- >./gtest
- 2) Project files, structure, classes and functions: DBFile is extended and contains following classes.
- a) Generic DBFile Class: This extended version of DBFile is designed to serve both purposes. Hence it can actually be used for sorted and unsorted both types of heap files. Generic DBFile Class is a generic virtual class which acts as parent class for both the normal heap and sorted DBFile classes. This class contains certain virtual methods which are implemented independently by both children classes. These virtual methods are not common among the children classes hence their implementation is unique in both of the children classes. This class contains protected variables for File, Page, Comparison Engine class objects, it also contains a pointer for preference file. File class objects hold the pages and page class objects hold the records to be written on the disk. There are certain public functions which are common to both of the children classes, these functions are GetPageLocationToWrite() ,GetPageLocationToRead(BufferMode mode) , GetPageLocationToReWrite() , Create (char \* f\_path,fType f\_type, void \*startup) and Open (char \* f\_path). There are other virtual methods which are implemented independently by the children classes these methods are MoveFirst (), Add (), Load (), GenNext (), and Close () methods.

```
class GenericDBFile{
    File fileObj;
    Page pageObj;
    Preference * prefPointer;
    ComparisonEngine comparisonEngineObj;
   GenericDBFile();
   int GetPageLocationToWrite();
    int GetPageLocationToRead(BufferMode mode);
    int GetPageLocationToReWrite();
    void Create (char * f_path,fType fileType, void *startup);
    int Open (char * f_path);
   virtual ~GenericDBFile();
   virtual void MoveFirst ();
    virtual void Add (Record &addme);
    virtual void Load (Schema &myschema, const char *loadpath);
    virtual int GetNext (Record &fetchme);
    virtual int GetNext (Record &fetchme, CNF &cnf, Record &literal);
    virtual int Close();
};
```

b) HeapDBFile Class: This class Implements the DBFile unsorted Heap functionality implemented in project. This class contains all the relevant methods which were used to implement unsorted heaps. This class contains a constructor, destructor apart from that There are other virtual methods which are implemented independently by this class these methods are MoveFirst (), Add (), Load (), GenNext (), and Close () methods.

```
//This is Heap DB FIle Class which inherits and overrides virtual functions from GenericDBFile class
class HeapDBFile: public virtual GenericDBFile{
public:
    HeapDBFile(Preference * preference);
    ~HeapDBFile();
    void MoveFirst ();
    void Add (Record &addme);
    void Load (Schema &myschema, const char *loadpath);
    int GetNext (Record &fetchme);
    int GetNext (Record &fetchme, CNF &cnf, Record &literal);
    int Close ();
};
```

c) **SortedDBFile Class:** This Class contains all the methods which implement the new functionality of sorted heap files. SortedDBFile Class contains attributes such as input and output pipes which can push and retrieve records from the BlgQ class. This class also contains a pointer for BigQ class as well as File object which holds the pages. A buffer attribute is also used in this class which is used before the pages are written into file object. A queryorder maker attribute is used to construct a sort order with given CNF and also sort order of DBFile.

```
class SortedDBFile: public GenericDBFile{
    Pipe * inputPipePtr;
    Pipe * outputPipePtr;
    BigQ * bigQPtr;
    File newFile;
    Page outputBufferForNewFile;
    OrderMaker * queryOrderMaker;
    bool doBinarySearch;
public:
    SortedDBFile(Preference * preference);
   ~SortedDBFile();
    void MoveFirst ();
    void Add (Record &addme);
    void Load (Schema &myschema, const char *loadpath);
    int GetNext (Record &fetchme);
    int GetNext (Record &fetchme, CNF &cnf, Record &literal);
    int Close ();
    void MergeSortedInputWithFile();
    int BinarySearch(Record &fetchme, Record &literal, off_t low, off_t high);
};
```

### SortedDBFile Class Variables:

- 1) Pipe \* inputPipePtr: This pointer corresponds to a Pipe object. This is an input pipe which is used to write records into BigQ class objects.
- 2) Pipe \* outputPipePtr: This pointer corresponds to a Pipe object. This is an output pipe which is used for retrieving records from BigQ class objects.
- 3) BigQ \* bigQPtr: This pointer corresponds to BigQ class object. This BigQ class object is used to sort records which are loaded into sorted DBFile Object.
- **4) File newFile:** This variable corresponds to file object. A file object is used to store pages. Pages contains records. This file object is specifically used to merge inputs when the mode is changed from write to read.

- 5) Page outputBufferForNewFile: As mentioned above the file object in class is specifically used to merge inputs when the mode is changed from write to read. This outputBufferForNewFile variable is a pointer for a Page object which acts as a buffer for writing into file object which is being used to merge inputs. Ultimately this becomes the new sorted DB File in itself.
- 6) OrderMaker \* queryOrderMaker: This variable corresponds to order maker object. This query order make is used by GenNext method. This is used to construct a sort order for given CNF and also sort order of DBFile.

## 7) bool doBinarySearch:

## **SortedDBFile Class Functions:**

### 1) void MoveFirst ()

**Description:** MoveFirst () method implementation in DBFile is same as of heap file. In the heap file this function Forces the pointer to reset its position to the first record of the file. Each DBfile constitutes a pointer which points to the first record in the file. This pointer can move with respect to retrievals of records from the file. In a situation where the the page buffer is not empty and the DBFIle is in write mode .Then in that case the non-empty page buffer is first written into dbfile, the first page is read and the first record from the page is fetched.in other conditions the pointer is moved towards first page and it is read.

### 2) void Add (Record &addme)

**Description**: The class contains an attribute which is an input pipe which is used to write records into BigQ class objects. This Add function is used to write records in the input pipe. Incase if the buffer is set to read mode then before the records are written into pipe the buffer is set to the write mode. Once the buffer is set to write mode then sorted records form the output pipe of BigQ class objects are written into input pipe.

## 3) void Load (Schema &myschema, const char \*loadpath)

**Description:** The load function can add records into input pipe in bulk. Same as add function load function implements the same functionality the only difference is that the load function incorporates a loop to add records in bulk over add function which can only add one record at a time. the loop sucks records from the raw.tbl file, and till all records are not written load function while loop keeps on running.

#### 4) int GetNext (Record &fetchme)

**Description:** This method GetNext() gets records from the dbfiles iteratively. The persistent dbfile (.bin file) contains various pages which in turn contain records. the GetNext() method reads the pages present in the persistant file one by one and fetches all the records present in these pages iteratively when all the records are read from the file then this function returns 0.

#### 5) int GetNext (Record &fetchme, CNF &cnf, Record &literal)

**Description:** This version of GetNext method takes three arguments. One of the arguments is the CNF predicate. using a function like GetSortOrders of CNF class this given CNF is converted to an OrderMaker object. Then this given CNF is compared with the existing sorting order of dbfile. The attributes of ordermakers can be accommodated with the sort order of DBfile then in that case BinarySearch function is used to find any record that satisfy that CNF. We keep getting such records for each of the subsequent calls to this function.

### 6) int Close ()

**Description:** This function Close () simply closes the dbfile (.bin file). when the program has completely written the state of dbfile in a .pref file (preference file). the order maker object is also written in the state in the case of sorted DB File. it is done so that the sort order of the file could be known when it is read.

## 7) void MergeSortedInputWithFile ()

**Description:** When the program switches from write mode to the read mode then this function is used. In this state when the mode is switched then the records wich are already present in the BigQ class object have to be merged with the already sorted file. This function is more of a merge step which merges two sorted sources together. A temporary file is used for writing the merge stream.this temporary file is renamed to original dbfile.

## 8) int BinarySearch(Record &fetchme, Record &literal,off\_t low, off\_t high)

**Description:** This function is a function which is used in GetNext() method which also accepts a CNF as an argument. Binary search is applied on the pages of DBFile. The given CNF is compared with the existing sorting order of dbfile. The attributes of ordermakers can be accommodated with the sort order of DBfile then in that case BinarySearch function is used to find any record that satisfy that CNF. This function returns a 0 if no record is found satisfying the CNF.

3) Program Run Results: All the screenshots are added in screenshots folder of this submission.

## a) Test Case 1:

```
[azureuser113@VM-02:~/final/a2-2test$ ./test.out
** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
                        catalog
 catalog location:
 tpch files dir:
                        tpch-dbgen/
 heap files dir:
                        dbfiles/
 select test option:
         1. create sorted dbfile
         2. scan a dbfile
         3. run some query
 select table:
         1. nation
         2. region
         3. customer
         4. part
         5. partsupp
         6. supplier
         7. orders
         8. lineitem
 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
         added 150000 recs..so far 150000
 create finished.. 150000 recs inserted
azureuser113@VM-02:~/final/a2-2test$
```

### b) Test Case 2:

```
* IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
 catalog location:
tpch files dir:
heap files dir:
                                                                           dbfiles/
   select test option:
1. create sorted dbfile
2. scan a dbfile
                                     run some query
  select table:
                                   nation
  scan : dbfiles/customer.bin
  c_custey: [5998], c_name: [Customer#80005998], c_address: [Ddy7dHn EQ1], c_nationkey: [0], c_phone: [10-100-100-117], c_acctbal: [5956.88], c_mktsegment: [MuCHINERY], c_comment: [, even epitaphs wake furiously permanently pending frays. evenly pending accounts capicle carefully. pending pack]
c_custeys: [4254], c_mmer: [Customer#800042145], c_address: [HFG56HeXtTnQP0ZAODStr], c_nationkey: [0], c_phone: [10-100-274-5785], c_acctbal: [7128], c_mktsegment: [FURNITURE], c_comment: [ts. Tiresias across the quickly regular f]
c_custeys: [42556], c_mmer: [Customer#800006596], c_address: [n9780015MKNIaAzuCVw17gg16kCpiO8r,Rj15F9e], c_nationkey: [0], c_phone: [10-100-274-5785], c_acctbal: [6378.16], c_mktsegment: [HOUSEHOLD], c_comment: [thely final accounts haggle slyly. theodolites integrate al
 .c_custkey: [5994], c_name: [Customer#000367964], c_address: [QAMG5/28iyjq9U4IS], c_nationkey: [0], c_phone: [10-100-301-2651], c_acctbal: [6312.45], c_nktsegment: [MACHINERY], c_comment: [even accounts. unusual foxes wake according to the unusual frays. bli]

c_custkey: [59984], c_name: [Customer#00036984], c_address: [6xYqTYQ0rRmFY71rRTh], c_nationkey: [0], c_phone: [10-100-541-5585], c_acctbal: [6405.08], c_mktsegment: [MACHINERY], c_comment: [even sauternes haggle according to the furiously regular deposits. requests slee
      _custkey: [189996], c_name: [Customer#800199996], c_address: [pACgw87b6H/pul], c_nationkey: [0], c_phone: [18-180-581-8579], c_acctbal: [-611.47], c_mktsegment: [HOUSEHOLD], c_comment: [y even foxes boost carefully carefully ironic packages bold packages about the blith
                nusual opennes
stkey: [31573], c_name: [Customer#000031757], c_address: [fy7,76n]jkA39R73044f194]jGBrQ0cWi], c_nationkey: [0], c_phone: [10-100-588-9355], c_acctbal: [7731.01], c_mktsegment: [FURNITURE], c_comment: [r deposits? pending ideas sleep furiously. requests ]
stkey: [81501], c_name: [Customer#0000031691], c_address: [kXDhS,]PmMgd,mkVk], c_nationkey: [0], c_phone: [10-100-864-5796], c_acctbal: [5188.79], c_mktsegment: [BUILDING], c_comment: [sual packages deposits along the final pinto beans doze stealthily along]
stkey: [1073331], c_name: [Customer#000107333], c_address: [UKH5g9XgF8YIv6j6i70eX0y1zmx3FF], c_nationkey: [0], c_phone: [10-101-152-4542], c_acctbal: [674.96], c_mktsegment: [MACHINERY], c_comment: [ to the final requests. slyly regular packages at the furiously even
c_custkey: [187333], c_name: [Customer#800873333], c_address: [UNGSFVRFSVIv6j4170avVylrmWSFF], c_nationkey: [0], c_phone: [18-101-1527-4547], c_actibal: [6/4.795], c_mktsegment: [MACHINERY], c_comment: [to the rinal requests styly regular parkages at the following parkages at t
   azungszue uj
c_custkey: [3339], c_name: [Customer#000083339], c_address: [CyvgNvTJpEy8N8 yNzM336KY.72528Q8Nvou], c_nationkey: [0], c_phone: [10-102-409-7401], c_acctbal: [7724.95], c_mktsegment: [BUILDING], c_comment: [n asymptotes. bold instructions sleep carefully. regular excuse]
c_custkey: [80836], c_name: [Customer#000080836], c_address: [CJv9RX4Tp18i 1470m66in,2ceQJ8K,], c_mationkey: [0], c_phone: [10-102-549-6244], c_acctbal: [-77.68], c_mktsegment: [AUTOMOBILE], c_comment: [ose furiously ironic accounts. ironic deposits are blithely. ironic
  c. custkey: [64573], c_name: [Customer#800064573], c_address: [10tbtjAen3c yR8], c_nationkey: [0], c_phone: [10-102-554-6504], c_acctbal: [5705.04], c_mktsegment: [AUTOMOBILE], c_comment: [efully ironic braids are. furiously regular deposits use fluffily unusual p]
                           yy: [93747], c_name: [Customer#0000993747], c_address: [uoDW0hEIzEl1KCxbL2T9kERIjmDpE]wZk9Vx1U], c_nationkey: [24], c_phone: [34-992-958-7984], c_acctbal: [9455.22], c_mktsegment: [FURNITURE], c_comment: [ unusual theodolites doze. blithely final accounts along thi
                                  \tage1 (113698), c_name: [Customer#000113698], c_address: [1:UMBChd70eggr%SQC6x46371WXT8gff6 687], c_nationkey: [24], c_phone: [34-993-166-5163], c_acttbal: [-486,38], c_mktsegment: [HOUSEHOLD], c_comment: [onic ideas are. platelets could have to sleep efte] [126988], c_name: [Customer#00013698], c_address: [bd6mhyN8u,1], c_nationkey: [24], c_phone: [34-993-26-3601], c_acttbal: [756.4.2], c_aktsegment: [MUNINEY], c_comment: [y final instructions about the newer even accounts print carefully special reques] [27893], c_name: [customer#00027093], c_address: [GCC770mG674Ach)Printsyrighter), pfer the reg packages across the unusual instructions use slyly after the reg packages across the unusual instructions use slyly after the reg packages across the unusual instructions use slyly after the reg package across the unusual instructions use slyl after the reg package across the unusual instructions use slyl after the reg package across the unusual instructions use slyl after the reg package across the unusual instructions use slyl after the reg package across the unusual instructions use slyl after the reg package across the unusual instructions used slyl after the reg package across the unusual instructions used slyl after the reg package across the unusual instructions used slyl across the unusual instructions across the unusual 
                                   (42), c_name: [Customer#800844211], c_address: [glh5iz8BID,9], c_nationkey: [24], c_phone: [34-993-926-9853], c_acctbal: [637.97], c_mktsegment: [MACHINERY], c_comment: [. fluffily special patterns are about the evenly regular grouches. carefully special
                  tkey: [43757], _name: [Customer#00004577], c_address: [51ef55RYEYXIF5JAK8], c_nationkey: [24], c_phone: [34-994-145-7928], c_actbal: [-929.94], c_mktsegment: [MACHINERY], c_comment: [ironic frets wake regularly even asymptotes. regular ideas haggle against the inding diagnats; carefully f]

**tay: [13881], c_name: [Customer#00013801], c_address: [15070557770] HWX10792Dc45804], c_nationkey: [24], c_phone: [34-994-40-9043], c_actbal: [5935.36], c_mtsegment: [MACHINERY], c_comment: [cick] ironic excuses wake. unusual, final deposit]

**tay: [13881], c_name: [Customer#0000890], c_address: [Swr12xxx09x0808] 99;2xwl.[Ltris000277], c_nationkey: [24], c_phone: [34-994-809], c_actbal: [1871.86], c_mtsegment: [MULTONIC], c_comment: [ng, silent forces try to cipile act

**tay: [128494], c_name: [Customer#000028040], c_address: [wildpaids], c_nationkey: [24], c_phone: [34-994-803-4080], c_actbal: [401.86], c_mtsegment: [BULTONIC], c_comment: [24], special forces steps. carefully unusual requests according to the

**tay: [128491], c_name: [Customer#000080050], c_address: [tysiXNigiNTwr8NTpmm QRin], c_nationkey: [24], c_phone: [34-994-893-40724], c_actbal: [401.80], c_mtsegment: [BULTONIC], c_comment: [24], special forces steps. carefully unusual requests according to the

**tay: [40840], c_name: [401.80], c_address: [401.80
        custivy [19746], _name: [Oustomer#000019760], __address: [R05]iRgUUjpTPHx5K302vcZpJOX,], _nationkey: [24], __phone: [34-995-194-4009], __acctbal: [5286.51], __nktsegment: [AUTOMOBILE], __comment: [nusual requests cajole even, even packages. platelets are slyly. flut
                                     deposits]
[57978], __name: [Customer#000057078], c_address: [PqlN0uNmiMntx70ha], __nationkey: [24], c_phone: (34-995-249-6002], c_acctbal: [3861.85], c_mktsegment: [FURNITURE], c_comment: [al epitaphs detect furiously across the regular packages. slyly ironic ]
[66156], c_mane: [Customer#000057078], c_address: [,YgacKYlresUSidygulE346MuAnX80yLTuS], c_nationkey: [24], c_phone: (34-995-978-2279), c_acctbal: [5186.45], c_mtsegment: [FURNITURE], c_comment: [furiously pending theodolites sleep after th]
[74263], c_name: [Customer#000074263], c_address: [88kuJXq44binX50nJ,50t6 ,ommrxsDyGVY0Jf], c_nationkey: [24], c_phone: [34-995-950-2170], c_acctbal: [7803.79], c_mtsegment: [MACHINERY], c_comment: [unts impress. slyly final foxes haggle carefully around the comment of the comment 
                        ey: [88858], c_name: [Customer#800088858], c_address: [1Mshbshap]], c_nationkey: [24], c_phone: [34-996-665-1799], c_acctbal: [6866.8], c_mktsegment: [AUTOMOBILE], c_comment: [theodolites after the fluffily pending theodolites boost bold accounts. furiously repular [asjoin repular]
eye; [139443], c_name: [Customer#800139443], c_address: [26MgKikBcs, SP LtunUmSVV30,bvyaJv2HBOM], c_nationkey: [24], c_phone: [34-996-663-4478], c_acctbal: [-395.28], c_sktsegment: [BUILDING], c_comment: [ollphins integrate blithely carefully pending pinto beans. per
                        .ey: [37326], c_name: [Customer#000057326], c_address: [00X3cxIC7H29Vd5klTHchRdxHDpfUUS], c_nationkey: [24], c_phone: [34-996-655-7639], c_acctbal: [3155.24], c_mktsegment: [HOUSEHOLD], c_comment: [inal accounts sleep slyly regular pains. c]
ey: [25965], c_name: [Customer#000025985], c_address: [m wkl00 AZTfsAYDRct0ZDAoUe], c_nationkey: [24], c_phone: [34-996-827-5819], c_acctbal: [6393.65], c_mktsegment: [BUILDING], c_comment: [sly across the blithely even deposits. slyly pending excuses nag slyly. r
                             y: [6048], c_name: [Customer#800086048], c_address: [ONTLKGdov9MnFfAGa40Pidc6YIr5j 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. accounts wake. slyly fits above th]
                                y: [35264), __name: [Cuttomer#000001655), __address: [0 MAGOVERIBPDSPSP, __nationkey: [24], c_phone: [34-997-517-4668], [238-1], __nationkey: [24], c_phone: [34-997-617-4668], c_actrbal: [4429.09], __mktsegment: [AUTOMOBILE], c_comment: [cuttoms cajole ideas. slyly express forges use after the carefully ironic ideas. slyly express forges use after the carefully ironic ideas. slyly express forges use after the carefully ironic ideas. slyly express forges use after the carefully ironic ideas. slyly express forges use after the carefully ironic ideas. slyly express forges use after the carefully repetal [470-40], __address: [actro-fire/pakt2e3], __address
                                                                      Chame: [Customer#898182589], c_address: [uPFxngvS3aDQIL], c_nationkey: [24], c_phone: [34-999-363-7145], c_acutbal: [6256.97], c_mktsegment: [AUTOMOBILE], c_comment: [ly regular asymptotes haggle fluffily according to the furiously bold courts. dar
    . [austwop: [36478], c_name: [Customer#000035678], c_address: [OffScryilininXxivXvYZeFoOlth.iovs6cy], c_nationkey: [24], c_phone: [34-999-379-6813], c_actclast: [AUTOMOBILE], C_comment: [1y regular asymptotes haggle fluffly according to the furiously bold courts. dari castewy: [193586], c_name: [Customer#000018566], c_address: [CHSCNIGHEN], c_comment: [Ithout the fluffly ironic packages. daring, even del castewy: [193586], c_name: [Customer#000018566], c_address: [CHSCNIGHEN], c_comment: [Ithout the fluffly ironic packages. daring, even del castewy: [193586], c_nationic packages. daring, even del castewy: [193586], c_nationic packages. [Ithout the fluffly ironic packages. daring, even del castewy: [19378], c_name: [actclast: [Ithout the fluffly] even deposits boost. quickly ironic requests could have to castewy: [19378], c_name: [actclast: [1937], c_name: [actclast: [1937]
   scanned 150000 recs
```

#### c) Test Case 3:

```
## DOCATALT MAKE SAME THE INFORMATION RECOWNS CORRECT **

** PROCEDURE THE DESCRIPTION RECOUNT THE DESCRIPTION RECOWNS CORRECT **

** PROCEDURE THE DESCRIPTION RECOWNS CORRECT **

** PROCEDURE THE DESCRIPTION RECOWNS CORRECT **

** PROCEDURE THE DESCRIPTION RECOWNS CORRECT **

** P
```

#### d) Gtest Results: