Database System Implementation – Project 1
(Heap File Implementation)

Team Members:

Ananda Bhaasita Desiraju
(UFID: 4081-1191)

Nidhi Shashi Sharma
(UFID: 6843-1215)

Instructions to compile and run the code:

- To compile the code: make
- To compile the tests in Test.cc: make test.out
- To run the tests in Test.cc: ./test.out
- To compile the tests in GTests.cc: make GTests.out
- To run all the GTests in GTests.cc: ./GTests.out

Heap File Implementation:

Variables used in DBFile class:

- `file`:
   A pointer to the file used for processing
- `readPage, writePage`:
   Pointers to the page which is currently read or written to.
- `currentRecord`:
   A pointer referencing the current record to be accessed.
- `pageOffset, writeOffset`:
   Offset to the page in the file or point in the page from where data is read or written to.
- `dirtyWrite`:
   A flag that indicates switch between read and writes.
- `Eof`:
   A flag indicating the reaching of end of file.

Methods Implementation:

- `DBFile()`:
   Constructor of the class DBFile. Creates instances of a File and Pages each for reading and writing data as well as a record.

- `~DBFile()`:
   Destructor of the class DBFile. Destroys the instances of file, pages, and records created while exiting the program

- `Create(const char *fpath, fType file_type, void *startup):`
  In this project, there is only the implementation of the heap file system. So, this function creates a new file and sets the page offset to the first page and the write offset to the beginning of the file. Otherwise, in general, this function is supposed to create a new file of the file type selected i.e. Heap, sorted or tree.

- `Open(const char *fpath):`
  This function makes the assumption that the DBFile indicated by the fpath variable already exists and it tries to open the file. It takes the path of the file to be opened, as its input parameter. If the file that is to be opened has already been created and is closed, the function returns 1, indicating the file open is successful. It will throw an error otherwise.

- `Close():`
  This function closes the file that is currently open. It returns a value greater than 0 when the closing of file has been successful and a negative value otherwise.

- `Load(Schema &myschema, const char *loadpath):`
  The Load function bulk loads the DBFile instance from a text file, appending new data to it using the SuckNextRecord function from Record.h. The character string passed to Load is the name of the data file to bulk load. In our implementation, we create an object of type record and while there is data in the SuckNextRecord, we append it to our created record.

- `MoveFirst():`
  Each DBFile instance has a "pointer" to the current record in the file. By default, this pointer is at the first record in the file, but it can move in response to record retrievals. This function forces the pointer to correspond to the first record in the file. While the file if not empty, the first page is fetched and stored in the readPage pointer, otherwise the readPage pointer is emptied.

- `Add(Record &addme):`
  This function is used to add records to the file and in the heap implementation, the records are simply added to the end of the file. It consumes addMe and it cannot be put in the file again after being consumed.

- `GetNext(Record &fetchme):`
  GetNext fetches the next record from the file and returns, where "next" is defined to be relative to the current location of the pointer, indicated by the page Offset. Returning the record, the offset is also incremented, so the next call to GetNext won't return the same record twice. The return value is an integer whose value is zero if there is no valid record to be returned and 1 otherwise.

- `GetNext(Record &fetchme, CNF &cnf, Record &literal):`
  This version of GetNext also accepts a selection predicate which is a conjunctive normal form expression and returns the next record in the file that is accepted by the selection predicate. The literal record is used to check the selection predicate and is created when the parse tree for the CNF is processed.
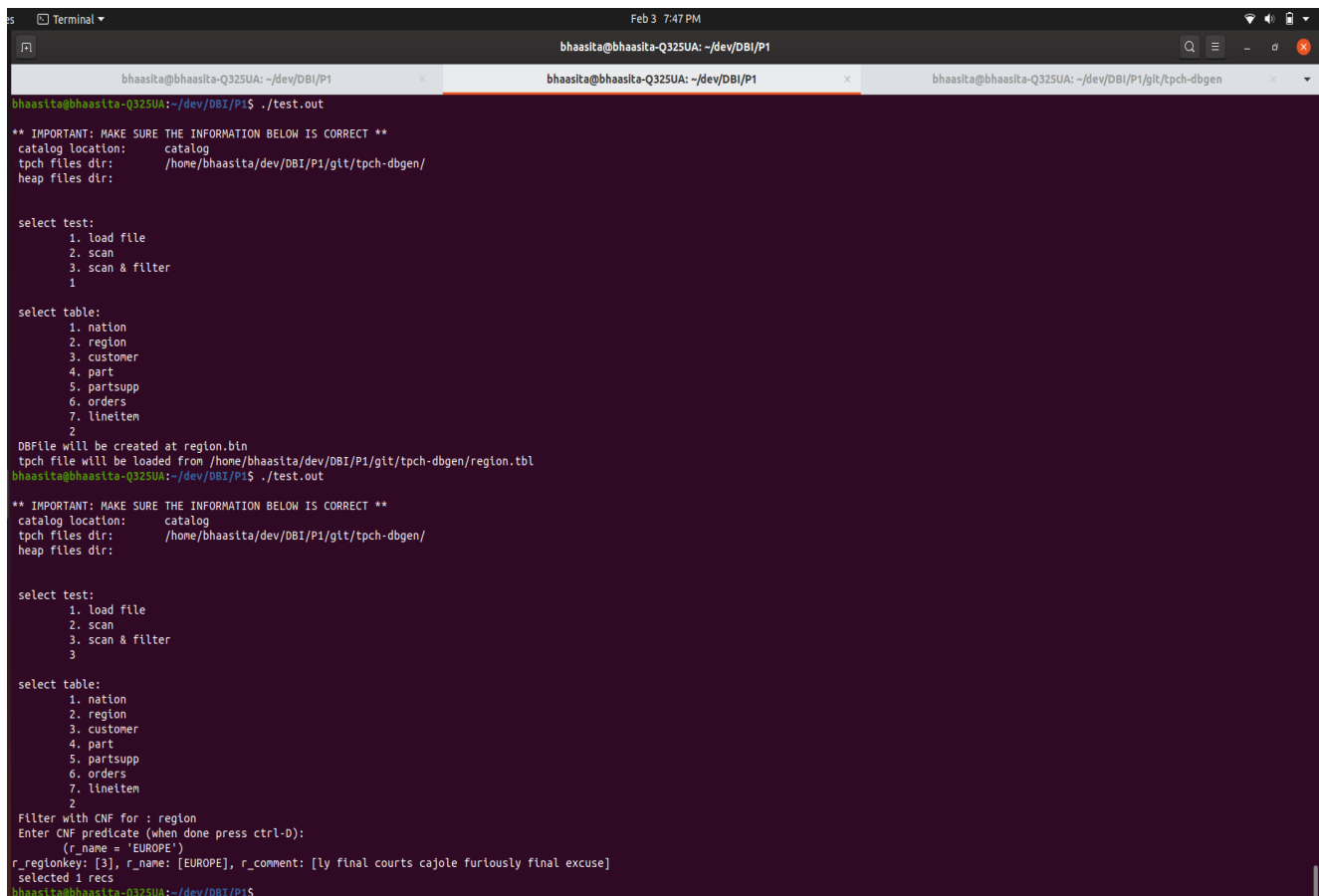
GTest Cases:

- CreateTest1:
  > This test case tests whether a DBFile has been successfully created or not.
- CreateTest2:
  > This test case tests whether an error is thrown in case of unsuccessful creation of a DBFile.
- OpenTest1:
  > This test case tests whether a DBFile has been successfully opened or not.
- OpenTest2:
  > This test case tests whether an error is thrown in case of unsuccessful opening of a DBFile.
- CloseFile:
- This test case tests a DBFile has been successfully closed.

SCREENSHOTS

1. For 10MB file:

- **Q1**

- **Q2**


```
bhaasita@bhaasita-Q325UA:~/dev/DBI/P1$ ./test.out

** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
 catalog location:        catalog
 tpch files dir:          /home/bhaasita/dev/DBI/P1/git/tpch-dbgen/
 heap files dir:

 select test:
         1. load file
         2. scan
         3. scan & filter
         3

 select table:
         1. nation
         2. region
         3. customer
         4. part
         5. partsupp
         6. orders
         7. lineitem
         2
 Filter with CNF for : region
 Enter CNF predicate (when done press ctrl-D):
         (r_name < 'middle east') AND
(r_regionkey > 1)
 r_regionkey: [2], r_name: [ASIA], r_comment: [ges. thinly even pinto beans ca]
 r_regionkey: [3], r_name: [EUROPE], r_comment: [ly final courts cajole furiously final excuse]
 r_regionkey: [4], r_name: [MIDDLE EAST], r_comment: [uickly special accounts cajole carefully blithely close requests
. carefully final asymptotes haggle furiousl]
  selected 3 recs
bhaasita@bhaasita-Q325UA:~/dev/DBI/P1$
```

Note: In case of region name = 'MIDDLE EAST' only two records are returned as the Ascii values of region name are being compared.


```
catalog location:        catalog
tpch files dir:          /home/bhaasita/dev/DBI/P1/git/tpch-dbgen/
heap files dir:

select test:
        1. load file
        2. scan
        3. scan & filter
        1

select table:
        1. nation
        2. region
        3. customer
        4. part
        5. partsupp
        6. orders
        7. lineitem
        2
DBFile will be created at region.bin
 tpch file will be loaded from /home/bhaasita/dev/DBI/P1/git/tpch-dbgen/region.tbl
bhaasita@bhaasita-Q325UA:~/dev/DBI/P1$ ./test.out

** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
 catalog location:       catalog
 tpch files dir:         /home/bhaasita/dev/DBI/P1/git/tpch-dbgen/
 heap files dir:

 select test:
         1. load file
         2. scan
         3. scan & filter
         3

 select table:
         1. nation
         2. region
         3. customer
         4. part
         5. partsupp
         6. orders
         7. lineitem
         2
 Filter with CNF for : region
 Enter CNF predicate (when done press ctrl-D):
         (r_name < 'MIDDLE EAST') AND
(r_regionkey > 1)

 r_regionkey: [2], r_name: [ASIA], r_comment: [ges. thinly even pinto beans ca]
 r_regionkey: [3], r_name: [EUROPE], r_comment: [ly final courts cajole furiously final excuse]
  selected 2 recs
bhaasita@bhaasita-Q325UA:~/dev/DBI/P1$
```

- **Q3(where n_name = 'japan')**


```
** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
 catalog location:       catalog
 tpch files dir:         /home/bhaasita/dev/DBI/P1/git/tpch-dbgen/
 heap files dir:

 select test:
        1. load file
        2. scan
        3. scan & filter
        1

 select table:
        1. nation
        2. region
        3. customer
        4. part
        5. partsupp
        6. orders
        7. lineitem
        1
 DBFile will be created at nation.bin
 tpch file will be loaded from /home/bhaasita/dev/DBI/P1/git/tpch-dbgen/nation.tbl
bhaasita@bhaasita-Q325UA:~/dev/DBI/P1$ ./test.out
** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
 catalog location:       catalog
 tpch files dir:         /home/bhaasita/dev/DBI/P1/git/tpch-dbgen/
 heap files dir:

 select test:
        1. load file
        2. scan
        3. scan & filter
        3

 select table:
        1. nation
        2. region
        3. customer
        4. part
        5. partsupp
        6. orders
        7. lineitem
        1
 Filter with CNF for : nation
 Enter CNF predicate (when done press ctrl-D):
        (n_regionkey = 3) AND
(n_nationkey > 10) AND
(n_name > 'japan')
 selected 0 recs
bhaasita@bhaasita-Q325UA:~/dev/DBI/P1$
```

- **Q3(where n_name = 'JAPAN')**


```
bhaasita@bhaasita-Q325UA:~/dev/DBI/P1$ ./test.out

** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
 catalog location:       catalog
 tpch files dir:         /home/bhaasita/dev/DBI/P1/git/tpch-dbgen/
 heap files dir:

 select test:
        1. load file
        2. scan
        3. scan & filter
        3

 select table:
        1. nation
        2. region
        3. customer
        4. part
        5. partsupp
        6. orders
        7. lineitem
        1
 Filter with CNF for : nation
 Enter CNF predicate (when done press ctrl-D):
        (n_regionkey = 3) AND
(n_nationkey > 10) AND
(n_name > 'JAPAN')
n_nationkey: [19], n_name: [ROMANIA], n_regionkey: [3], n_comment: [ular asymptotes are about the furious multiplier
s. express dependencies nag above the ironically ironic account]
n_nationkey: [22], n_name: [RUSSIA], n_regionkey: [3], n_comment: [ requests against the platelets use never accordi
ng to the quickly regular pint]
n_nationkey: [23], n_name: [UNITED KINGDOM], n_regionkey: [3], n_comment: [eans boost carefully special requests. ac
counts are. carefull]
 selected 3 recs
bhaasita@bhaasita-Q325UA:~/dev/DBI/P1$
```

- **Q11 (Note: l_quantity = 4.0)**

- **Q12**

2. For 1GB file:

- **Q1**

- **Q2**

- **Q3(where n_name = 'japan')**


```
** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location:      catalog
tpch files dir:        /home/bhaasita/dev/DBI/P1/git/tpch-dbgen/
heap files dir:

select test:
        1. load file
        2. scan
        3. scan & filter
        1

select table:
        1. nation
        2. region
        3. customer
        4. part
        5. partsupp
        6. orders
        7. lineitem
        1
 DBFile will be created at nation.bin
 tpch file will be loaded from /home/bhaasita/dev/DBI/P1/git/tpch-dbgen/nation.tbl
bhaasita@bhaasita-Q325UA:~/dev/DBI/P1$ ./test.out
** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location:      catalog
tpch files dir:        /home/bhaasita/dev/DBI/P1/git/tpch-dbgen/
heap files dir:

select test:
        1. load file
        2. scan
        3. scan & filter
        3

select table:
        1. nation
        2. region
        3. customer
        4. part
        5. partsupp
        6. orders
        7. lineitem
        1
 Filter with CNF for : nation
 Enter CNF predicate (when done press ctrl-D):
        (n_regionkey = 3) AND
(n_nationkey > 10) AND
(n_name > 'japan')
 selected 0 recs
bhaasita@bhaasita-Q325UA:~/dev/DBI/P1$
```

- **Q3(where n_name = 'JAPAN'**


```
select test:
        1. load file
        2. scan
        3. scan & filter
        3

select table:
        1. nation
        2. region
        3. customer
        4. part
        5. partsupp
        6. orders
        7. lineitem
        1
 Filter with CNF for : nation
 Enter CNF predicate (when done press ctrl-D):
        (n_regionkey = 3) AND
(n_nationkey > 10) AND
(n_name > 'japan')
 selected 0 recs
bhaasita@bhaasita-Q325UA:~/dev/DBI/P1$ ./test.out

** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location:      catalog
tpch files dir:        /home/bhaasita/dev/DBI/P1/git/tpch-dbgen/
heap files dir:

select test:
        1. load file
        2. scan
        3. scan & filter
        3

select table:
        1. nation
        2. region
        3. customer
        4. part
        5. partsupp
        6. orders
        7. lineitem
        1
 Filter with CNF for : nation
 Enter CNF predicate (when done press ctrl-D):
        (n_regionkey = 3) AND
(n_nationkey > 10) AND
(n_name > 'JAPAN')
n_nationkey: [19], n_name: [ROMANIA], n_regionkey: [3], n_comment: [ular asymptotes are about the furious multipliers. express dependencies nag above the ironically ironic account]
n_nationkey: [22], n_name: [RUSSIA], n_regionkey: [3], n_comment: [ requests against the platelets use never according to the quickly regular pint]
n_nationkey: [23], n_name: [UNITED KINGDOM], n_regionkey: [3], n_comment: [eans boost carefully special requests. accounts are. carefull]
 selected 3 recs
bhaasita@bhaasita-Q325UA:~/dev/DBI/P1$
```

- **Q11**


```
catalog location:        catalog
tpch files dir:          /home/bhaasita/dev/DBI/P1/git/tpch-dbgen/
heap files dir:

select test:
        1. load file
        2. scan
        3. scan & filter
        1

select table:
        1. nation
        2. region
        3. customer
        4. part
        5. partsupp
        6. orders
        7. lineitem
        7
DBFile will be created at lineitem.bin
tpch file will be loaded from /home/bhaasita/dev/DBI/P1/git/tpch-dbgen/lineitem.tbl
bhaasita@bhaasita-Q325UA:~/dev/DBI/P1$ ./test.out

** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location:        catalog
tpch files dir:          /home/bhaasita/dev/DBI/P1/git/tpch-dbgen/
heap files dir:

select test:
        1. load file
        2. scan
        3. scan & filter
        3

select table:
        1. nation
        2. region
        3. customer
        4. part
        5. partsupp
        6. orders
        7. lineitem
        7
Filter with CNF for : lineitem
Enter CNF predicate (when done press ctrl-D):
        (l_shipdate > '1994-01-01') AND
(l_shipdate < '1994-01-07') AND
(l_discount > 0.05) AND
(l_discount < 0.06) AND
(l_quantity = 4.0)
selected 0 recs
bhaasita@bhaasita-Q325UA:~/dev/DBI/P1$
```

- **Q12**


```
bhaasita@bhaasita-Q325UA:~/dev/DBI/P1$ ./test.out

** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location:        catalog
tpch files dir:          /home/bhaasita/dev/DBI/P1/git/tpch-dbgen/
heap files dir:

select test:
        1. load file
        2. scan
        3. scan & filter
        3

select table:
        1. nation
        2. region
        3. customer
        4. part
        5. partsupp
        6. orders
        7. lineitem
        7
Filter with CNF for : lineitem
Enter CNF predicate (when done press ctrl-D):
        (l_orderkey > 100) AND
(l_orderkey < 1000) AND
(l_partkey > 100) AND
(l_partkey < 5000) AND
(l_shipmode = 'AIR') AND
(l_linestatus = 'F') AND
(l_tax < 0.07)
l_orderkey: [130], l_partkey: [1739], l_suppkey: [4240], l_linenumber: [2], l_quantity: [48], l_extendedprice: [78755], l
_discount: [0.03], l_tax: [0.02], l_returnflag: [R], l_linestatus: [F], l_shipdate: [1992-07-01], l_commitdate: [1992-07-
12], l_receiptdate: [1992-07-24], l_shipinstruct: [NONE], l_shipmode: [AIR], l_comment: [lithely alongside of the regu]
l_orderkey: [194], l_partkey: [2594], l_suppkey: [5095], l_linenumber: [1], l_quantity: [17], l_extendedprice: [25442], l
_discount: [0.05], l_tax: [0.04], l_returnflag: [R], l_linestatus: [F], l_shipdate: [1992-05-24], l_commitdate: [1992-05-
22], l_receiptdate: [1992-05-30], l_shipinstruct: [COLLECT COD], l_shipmode: [AIR], l_comment: [ regular deposi]
selected 2 recs
bhaasita@bhaasita-Q325UA:~/dev/DBI/P1$
```

3. GTests:

```
[==========] Running 5 tests from 5 test cases.
[----------] Global test environment set-up.
[----------] 1 test from DBFileTest1
[ RUN      ] DBFileTest1.CreateTest1
[       OK ] DBFileTest1.CreateTest1 (8 ms)
[----------] 1 test from DBFileTest1 (8 ms total)

[----------] 1 test from DBFileTest2
[ RUN      ] DBFileTest2.CreateTest2
[       OK ] DBFileTest2.CreateTest2 (3 ms)
[----------] 1 test from DBFileTest2 (3 ms total)

[----------] 1 test from DBFileTest3
[ RUN      ] DBFileTest3.OpenTest1
[       OK ] DBFileTest3.OpenTest1 (0 ms)
[----------] 1 test from DBFileTest3 (0 ms total)

[----------] 1 test from DBFileTest4
[ RUN      ] DBFileTest4.OpenTest2
[       OK ] DBFileTest4.OpenTest2 (3 ms)
[----------] 1 test from DBFileTest4 (3 ms total)

[----------] 1 test from DBFileTest5
[ RUN      ] DBFileTest5.CloseFile
[       OK ] DBFileTest5.CloseFile (397 ms)
[----------] 1 test from DBFileTest5 (397 ms total)

[----------] Global test environment tear-down
[==========] 5 tests from 5 test cases ran. (412 ms total)
[  PASSED  ] 5 tests.
bhaasita@bhaasita-Q325UA:~/dev/DBI/P1$
```