

PROJECT REPORT

Group members:

Name- Shubham Saoji

Ufid- 26364957

How to run project:

make clean //to remove compiled files

make //to compile all files

make test.out //to compile test.out

//Enter appropriate options

./test [1-6] //to run query, specify query to be run

make gtest //to compile google test cases

./gtest //to run test cases

Queries:

1.

```
select *
from partsupp
where ps_supplycost < 1.03;
```
2.

```
select p_partkey, p_name, p_retailprice
from part
where (p_retailprice > 931.01) AND (p_retailprice < 931.3);
```
3.

```
select sum (s_acctbal + (s_acctbal * 1.05))
from supplier;
```
4.

```
select sum (ps_supplycost)
from supplier, partsupp
where s_suppkey = ps_suppkey;
```
5.

```
select distinct ps_suppkey
from partsupp
where ps_supplycost < 100.11;
```
6.

```
select sum (ps_supplycost) s_nationkey
from supplier, partsupp
where s_suppkey = ps_suppkey groupby s_nationkey;
```

BigQ Class and Methods:

BigQ Class consists of class members:

- in – For input pipe
- out – For output pipe

- sortorder – keeps sorting rule
- runlen – number of pages in each run

Other Helper Classes:

RecordTracker – to track record and pages

ThreadData – Passes pipe, sortorder and runlength data in threads

SortedFile Class and Methods:

nameOfFile: This variable stores the file whose path has been provided.

pipeIn: Instance of pipe class to refer input file

pipeOut: Instance of pipe class to refer input file

isDirty: true when currently working page is changed while doing writing operation; false when operation is done.

IndexOfRecord: stores index of record.

orderOfQuery: Instance of orderMaker which stores order of sorting.

currentRecord: stores the address of the record currently being used.

pIndex: stores int value of number of pages written in file (useful for writing purposes)

writeIndex: Indicates number of records that are written.

eofile: true when reached the end of the file; else false

SortedFile:

If the DBFile is specified to be sorted, we create instance of SortedFile. Methods in SortedFile class are:

Open:

For filetype sorted we call open method of SortedFile class. We pass in the OrderMaker value in this method. Opens the already created binary file and returns 1. If the file path is not present, it returns 0.

Create –

Creates a new empty database binary sorted file. We pass in parameter, startup having information regarding order of sorting and runlength. It returns 1 on file creation and 0 when no file path is given. After file creation, it records filepath of the binary file and initialized values to other members of the class.

MoveFirst:

Loads the first page of the binary file to memory and assigns ‘current’ pointer to the first record of the page.

Load:

loads all the records from tbl files to the respective database files; appends to the last. Handling: When all the records in tbl file are read completely, the records stored in the page in memory is written to the binary file, making sure all records are written to disk.

If the sorted DBFile is in “reading” mode, then the DBFile sets up its internal BigQ, adds the new records to it one-at-a-time via the queue’s input pipe, and then changes its mode to “writing”.

GetNext:

Fetch next record from file and return to user. If end of file is reached, return 0.

Add:

If we are Reading, we set `dirty_flag` to 1 and start appending the records from input to input pipe. We use `BigQ` object to use threading and sort.

RelOp:

SelectPipe, SelectFile, Project, Join, DuplicateRemoval, Sum, GroupBy, and WriteOut operations are encapsulated in this class

1. SelectFile

SelectFile takes a DBFile and a pipe as input. It also takes a CNF. It then performs a scan of the underlying file, and for every tuple accepted by the CNF, it stuffs the tuple into the pipe as output.

2. SelectPipe

SelectPipe takes two pipes as input: an input pipe and an output pipe. It also takes a CNF. It simply applies that CNF to every tuple that comes through the pipe, and every tuple that is accepted is stuffed into the output pipe.

3. Project

Project takes an input pipe and an output pipe as input. It also takes an array of integers `keepMe` as well as the number of attributes for the records coming through the input pipe and the number of attributes to keep from those input records. The array of integers tells Project which attributes to keep from the input records, and which order to put them in. So, for example, say that the array `keepMe` had the values [3, 5, 7, 1]. This means that Project should take the third attribute from every input record and treat it as the first attribute of those records that it puts into the output pipe. Project should take the fifth attribute from every input record and treat it as the second attribute of every record that it puts into the output pipe. The seventh input attribute becomes the third. And so on.

4. Join

Join takes two input pipes, an output pipe, and a CNF, and joins all of the records from the two pipes according to that CNF. Join should use a `BigQ` to store all of the tuples coming from the left input pipe, and a second `BigQ` for the right input pipe, and then perform a merge in order to join the two input pipes. You'll create the `OrderMakers` for the two `BigQ`'s using the CNF (the function `GetSortOrders` will be used to create the `OrderMakers`). If you can't get an appropriate pair of `OrderMakers` because the CNF can't be implemented using a sort-merge join (due to the fact it does not have an equality check) then your Join operation should default to a block-nested loops join.

5. DuplicateRemoval

DuplicateRemoval takes an input pipe, an output pipe, as well as the schema for the tuples coming through the input pipe, and does a duplicate removal. That is, everything that comes through the output pipe will be distinct. It will use the `BigQ` class to do the duplicate removal.

6. Sum

Sum computes the SUM SQL aggregate function over the input pipe, and puts a single tuple into the output pipe that has the sum.

7. GroupBy

GroupBy is a lot like Sum, except that it does grouping, and then puts one sum into the output pipe for each group. Every tuple put into the output pipe has a sum as the first attribute, followed by the values for each of the grouping attributes as the remainder of the attributes. The grouping is specified using an instance of the OrderMaker class that is passed in. The sum to compute is given in an instance of the Function class.

8. WriteOut

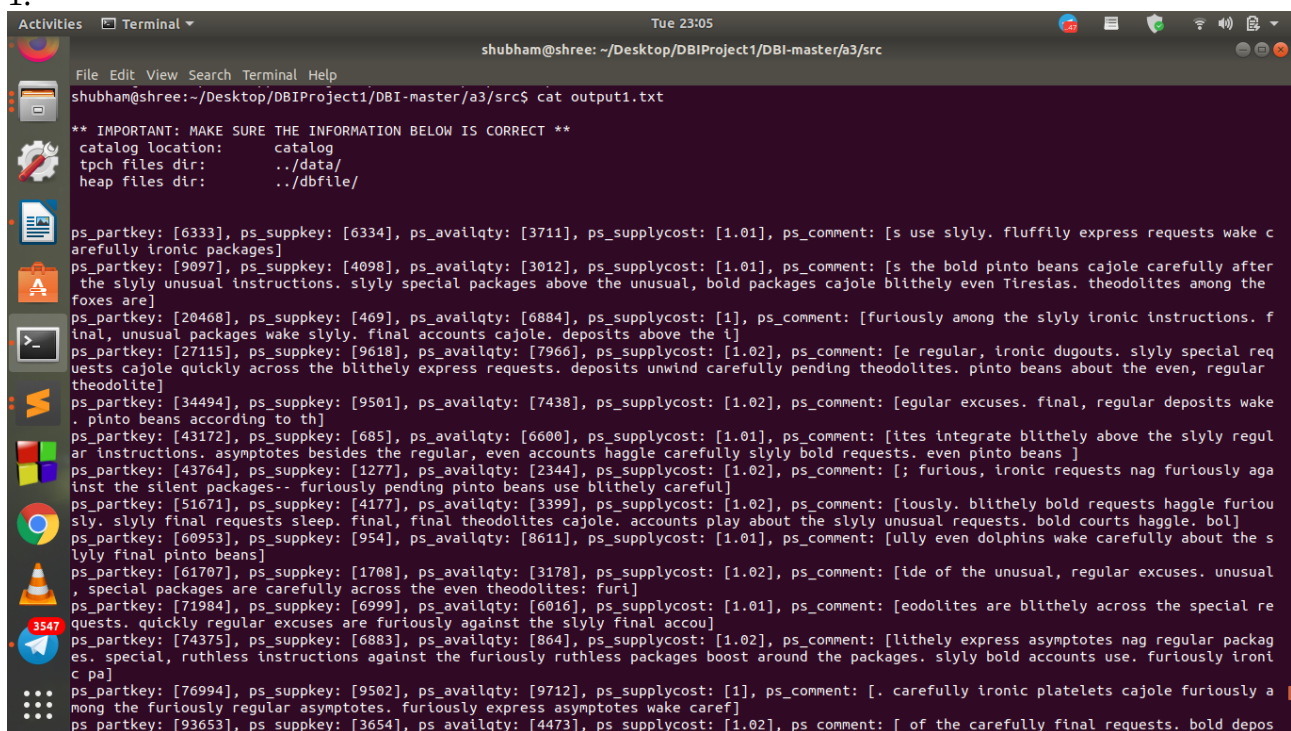
WriteOut accepts an input pipe, a schema, and a FILE*, and uses the schema to write text version of the output records to the file.

Output screenshots:

The output on running runTestCases.sh is stored in file output1.txt

Below are screenshot for output1.txt:

1.



```
shubham@shree: ~/Desktop/DBIProject1/DBI-master/a3/src$ cat output1.txt

** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location:      catalog
tpch files dir:        ../data/
heap files dir:        ../dbfile/

ps_partkey: [6333], ps_suppkey: [6334], ps_availqty: [3711], ps_supplycost: [1.01], ps_comment: [s use slyly. fluffily express requests wake c
arefully ironic packages]
ps_partkey: [9097], ps_suppkey: [4098], ps_availqty: [3012], ps_supplycost: [1.01], ps_comment: [s the bold pinto beans cajole carefully after
the slyly unusual instructions. slyly special packages above the unusual, bold packages cajole blithely even Tiresias. theodolites among the
foxes are]
ps_partkey: [20468], ps_suppkey: [469], ps_availqty: [6884], ps_supplycost: [1], ps_comment: [furiously among the slyly ironic instructions. f
inal, unusual packages wake slyly. final accounts cajole. deposits above the i]
ps_partkey: [27115], ps_suppkey: [9618], ps_availqty: [7966], ps_supplycost: [1.02], ps_comment: [e regular, ironic dugouts. slyly special req
uests cajole quickly across the blithely express requests. deposits unwind carefully pending theodolites. pinto beans about the even, regular
theodolite]
ps_partkey: [34494], ps_suppkey: [9501], ps_availqty: [7438], ps_supplycost: [1.02], ps_comment: [egular excuses. final, regular deposits wake
. pinto beans according to th]
ps_partkey: [43172], ps_suppkey: [685], ps_availqty: [6000], ps_supplycost: [1.01], ps_comment: [ites integrate blithely above the slyly regul
ar instructions. asymptotes besides the regular, even accounts haggle carefully slyly bold requests. even pinto beans ]
ps_partkey: [43764], ps_suppkey: [1277], ps_availqty: [2344], ps_supplycost: [1.02], ps_comment: [; furious, ironic requests nag furiously aga
inst the silent packages-- furiously pending pinto beans use blithely careful]
ps_partkey: [51671], ps_suppkey: [4177], ps_availqty: [3399], ps_supplycost: [1.02], ps_comment: [iously. blithely bold requests haggle furio
us. slyly final requests sleep. final, final theodolites cajole. accounts play about the slyly unusual requests. bold courts haggle. bol]
ps_partkey: [60953], ps_suppkey: [954], ps_availqty: [8611], ps_supplycost: [1.01], ps_comment: [ully even dolphins wake carefully about the s
lyly final pinto beans]
ps_partkey: [61707], ps_suppkey: [1708], ps_availqty: [3178], ps_supplycost: [1.02], ps_comment: [ide of the unusual, regular excuses. unusal
, special packages are carefully across the even theodolites: furi]
ps_partkey: [71984], ps_suppkey: [6999], ps_availqty: [6016], ps_supplycost: [1.01], ps_comment: [eodolites are blithely across the special re
quests. quickly regular excuses are furiously against the slyly final accou]
ps_partkey: [74375], ps_suppkey: [6883], ps_availqty: [864], ps_supplycost: [1.02], ps_comment: [lithely express asymptotes nag regular packag
es. special, ruthless instructions against the furiously ruthless packages boost around the packages. slyly bold accounts use. furiously ironi
c pa]
ps_partkey: [76994], ps_suppkey: [9502], ps_availqty: [9712], ps_supplycost: [1], ps_comment: [. carefully ironic platelets cajole furiously a
mong the furiously regular asymptotes. furiously express asymptotes wake caref]
ps_partkey: [93653], ps_suppkey: [3654], ps_availqty: [4473], ps_supplycost: [1.02], ps_comment: [ of the carefully final requests. bold depos
```

2.

```

Activities  Terminal  Tue 23:05
shubham@shree: ~/Desktop/DBIProject1/DBI-master/a3/src

File Edit View Search Terminal Help

plate]
ps_partkey: [193981], ps_supkey: [3982], ps_avallqty: [619], ps_supplycost: [1.01], ps_comment: [ic accounts after the unusual, regular instr
uctions grow carefully around the blithely unusual dependencies. pending accounts along the bl]

query1 returned 21 records

*****
*****
** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location:    catalog
tpch files dir:      ../data/
heap files dir:      ../dbfile/

>_

int: [31], string: [slate seashell steel medium moccasin], double: [931.03]
int: [1030], string: [orange floral olive ivory lace], double: [931.03]
int: [2029], string: [midnight brown dim violet almond], double: [931.02]
int: [3028], string: [puff slate tomato moccasin azure], double: [931.02]
int: [4027], string: [white ivory moccasin coral puff], double: [931.02]
int: [5026], string: [blanched blush pink light wheat], double: [931.02]
int: [6025], string: [purple medium light aquamarine dark], double: [931.02]
int: [7024], string: [forest rosy peach antique midnight], double: [931.02]
int: [8023], string: [mint salmon moccasin blanched beige], double: [931.02]
int: [9022], string: [peru misty sandy dark drab], double: [931.02]
int: [10021], string: [blush steel green sienna snow], double: [931.02]
int: [11020], string: [plum khaki powder beige peru], double: [931.02]

query2 returned 12 records

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

```

3.

```

Activities  Terminal  Tue 23:05
shubham@shree: ~/Desktop/DBIProject1/DBI-master/a3/src

File Edit View Search Terminal Help

query2 returned 12 records

*****
*****
** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location:    catalog
tpch files dir:      ../data/
heap files dir:      ../dbfile/

double: [9.24505e+07]

query3 returned 1 records

*****
*****
** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location:    catalog
tpch files dir:      ../data/
heap files dir:      ../dbfile/

query4
double: [4.00377e+08]
query4 returned 1 recs

*****
*****
** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location:    catalog
tpch files dir:      ../data/

```

4.

```
Activities Terminal Tue 23:05
shubham@shree: ~/Desktop/DBIProject1/DBI-master/a3/src

File Edit View Search Terminal Help
catalog location: catalog
tpch files dir: ../data/
heap files dir: ../dbfile/

double: [9.24505e+07]

query3 returned 1 records

*****
*****
** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location: catalog
tpch files dir: ../data/
heap files dir: ../dbfile/

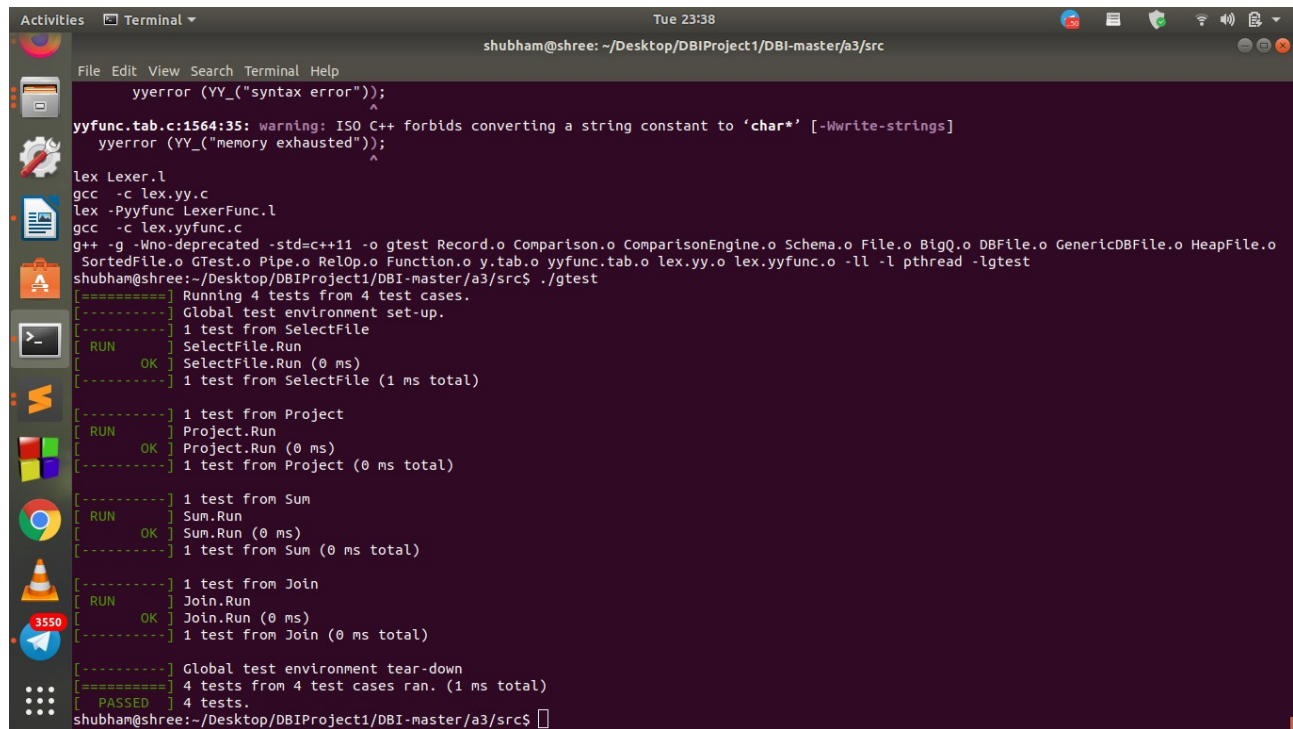
query4
double: [4.00377e+08]
query4 returned 1 recs

*****
*****
** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location: catalog
tpch files dir: ../data/
heap files dir: ../dbfile/

query5 finished..output written to file ps.w.tnp

shubham@shree:~/Desktop/DBIProject1/DBI-master/a3/src$
```

Screenshot of gtest:



```
shubham@shree: ~/Desktop/DBIProject1/DBI-master/a3/src
File Edit View Search Terminal Help
yyerror (YY_("syntax error"));
^
yyfunc.tab.c:1564:35: warning: ISO C++ forbids converting a string constant to 'char*' [-Wwrite-strings]
yyerror (YY_("memory exhausted"));
^
lex Lexer.l
gcc -c lex.yy.c
lex -Pyfunc LexerFunc.l
gcc -c lex.yyfunc.c
g++ -g -Wno-deprecated -std=c++11 -o gtest Record.o Comparison.o ComparisonEngine.o Schema.o File.o BigQ.o DBFile.o GenericDBFile.o HeapFile.o
SortedFile.o GTest.o Pipe.o RelOp.o Function.o y.tab.o yyfunc.tab.o lex.yy.o lex.yyfunc.o -ll -l pthread -lgtest
shubham@shree:~/Desktop/DBIProject1/DBI-master/a3/src$ ./gtest
[=====] Running 4 tests from 4 test cases.
[-----] Global test environment set-up.
[-----] 1 test from SelectFile
[ RUN      ] SelectFile.Run
[ OK       ] SelectFile.Run (0 ms)
[-----] 1 test from SelectFile (1 ms total)

[-----] 1 test from Project
[ RUN      ] Project.Run
[ OK       ] Project.Run (0 ms)
[-----] 1 test from Project (0 ms total)

[-----] 1 test from Sum
[ RUN      ] Sum.Run
[ OK       ] Sum.Run (0 ms)
[-----] 1 test from Sum (0 ms total)

[-----] 1 test from Join
[ RUN      ] Join.Run
[ OK       ] Join.Run (0 ms)
[-----] 1 test from Join (0 ms total)

[-----] Global test environment tear-down
[=====] 4 tests from 4 test cases ran. (1 ms total)
[ PASSED  ] 4 tests.
shubham@shree:~/Desktop/DBIProject1/DBI-master/a3/src$
```