

# **PROJECT REPORT**

# **RELATIONAL OPERATORS**

Name: Gloria Gilbert Nazareth

UFID: 8221-8035

GatorLink: [gnazareth@ufl.edu](mailto:gnazareth@ufl.edu)

**Steps to compile and execute:**

1. `make ./a2test.out` : this command builds the files required by `a2-test.cc` and creates the required object files
2. `./a2test.out` : This command will create all the bin files and will save it in the directory mentioned in `test.cat`
3. `make ./test.out` : this command builds the program and creates the required object files
4. `./test.out <input value 1-6>`: this will execute the program, and ask for user input in order to perform required functionality. Make sure `a2test.out` is executed before running it.
5. `./runTestCases.sh` : This command will execute all the test cases mentioned in the script file and will write the output to `output1.txt`
6. `make ./gtest.out` : this command will build the program and create the required object files
7. `./gtest.out` : this command will execute the gtests. Make sure `a2test.out` is executed before running gtests.

**Description of the Implemented Function:**

The `RelOp.cc` class has the following functions:

1. `void <ClassName>:: Run(<arguments>)`  
This function maps the arguments to the attributes of the class. It then creates a pthread to perform the required operation. Each of the class in `RelOp.cc` has a `Run` function which performs the same task.
2. `void <ClassName>::WaitUntilDone ()`  
This function joins the thread defined in the respective class, i.e it waits for the thread to complete its task. Each class defined in `RelOp.cc` has a `WaitUntilDone` function which performs the same task.
3. `void <ClassName>:: Use_n_Pages (int runlen)`  
This function assigns the value `runlen` to the `runLength` attribute defined in the respective class. All the classes in the `RelOp.cc` has a `Use_n_Pages` function which performs the same task.
4. `void* SelectFile::performOperation()`  
This function reads the data from the input file given to it, checks if it matches the CNF given. If it does, it inserts the record to the `outPipe`. If the CNF is null, it will insert all the records to the `outPipe`.
5. `void* SelectPipe::performOperation()`  
This function reads the data from the input pipe given to it, checks if it matches the CNF given. If it does, it inserts the record to the `outPipe`. If the CNF is null, it will insert all the records to the `outPipe`.

6. `void* Project::performOperation()`  
In this function records retrieved from the input pipe are projected to create new records containing columns which are mentioned in `keepMe` array. This new record with new schema is added to the `outPipe`.
7. `void* DuplicateRemoval::performOperation()`  
This function initially sorts all the records in its `inPipe` using `BigQ`. It then retrieves the records and compares adjacent records based on the schema given. If the record contains duplicate values, it inserts only one of them in the `outPipe`. Thus, the `outPipe` contains only unique values for schema value.
8. `void* Sum::performOperation()`  
This function retrieves the records from the `inPipe` and checks the data type of the column whose sum is to be calculated (whether `int` or `double`). It then adds the value to the total sum. Post this it creates a record using `Compose` function and adds the new record to the `outPipe`.
9. `void* GroupBy::performOperation()`  
This function performs grouping of records based on the grouping attributes and also calculates the sum for each group. It initially sorts all the records in the `inPipe` using `BigQ`. Post this it compares adjacent records and checks if they belong to the same group. Sum is calculated using the logic in point 8 for each group, and the new record for each group is composed of grouping attribute along with sum which is added to the `outPipe`.
10. `void* Join::performOperation()`  
This function joins two tables based on `CNF` given. Initially both the tables are sorted using `BigQ`. If `CNF` is given, we retrieve the records from the left and right table, and check if they match, else based on the values of the comparison we retrieve values from respective tables. Once the right and left have matched, we check if there are more records in right that match the same left record, and vice versa and add it in left and right vectors. Each record in each of the vector is merged with the other. If the `CNF` is not given all the records of the right table merged with that of the left (like cross product).
11. `void* WriteOut::performOperation()`  
This function retrieves the data from the input file and writes it to the `outFile` in the format of the data given in `tbl` files.

## Output on 1GB data:

### 1. Output of Query 1: [Snapshot of Output1.txt]

```
output1.txt (Work /media/gloria/Work1/UniversityOfFlorida/Study/Sem1/DB1/Projects/Project3/a3test) - gedit
Open  [F] Save

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

runLength is:::0
number of records inserted::21
ps_partkey: [6333], ps_suppkey: [6334], ps_avallqty: [3711], ps_supplycost: [1.01], ps_comment: [s use slyly. fluffly express requests wake carefully ironic packages]
ps_partkey: [9897], ps_suppkey: [4898], ps_avallqty: [3812], 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: [26468], ps_suppkey: [469], ps_avallqty: [6884], ps_supplycost: [1], ps_comment: [furiously among the slyly ironic instructions. final, unusual packages wake slyly. final accounts cajole.
deposits above the l]
ps_partkey: [27115], ps_suppkey: [9618], ps_avallqty: [7966], ps_supplycost: [1.02], ps_comment: [e regular, ironic dugouts. slyly special requests cajole quickly across the blithely express requests.
deposits unwind carefully pending theodolites. pinto beans about the even, regular theodolite]
ps_partkey: [34494], ps_suppkey: [9581], ps_avallqty: [7438], ps_supplycost: [1.02], ps_comment: [regular excuses. final, regular deposits wake. pinto beans according to th]
ps_partkey: [43172], ps_suppkey: [685], ps_avallqty: [6600], ps_supplycost: [1.01], ps_comment: [lites integrate blithely above the slyly regular instructions. asymptotes besides the regular, even accounts
haggle carefully slyly bold requests. even pinto beans ]
ps_partkey: [43764], ps_suppkey: [1277], ps_avallqty: [2344], ps_supplycost: [1.02], ps_comment: [; furious, ironic requests nag furiously against the silent packages-- furiously pending pinto beans use
blithely careful]
ps_partkey: [51671], ps_suppkey: [4177], ps_avallqty: [3399], ps_supplycost: [1.02], ps_comment: [lously. blithely bold requests haggle furiously. slyly final requests sleep. final, final theodolites
Cajole. accounts play about the slyly unusual requests. bold courts haggle. bol]
ps_partkey: [68953], ps_suppkey: [954], ps_avallqty: [8611], ps_supplycost: [1.01], ps_comment: [ully even dolphins wake carefully about the slyly final pinto beans]
ps_partkey: [61787], ps_suppkey: [1788], ps_avallqty: [3178], ps_supplycost: [1.02], ps_comment: [lde of the unusual, regular excuses. unusual, special packages are carefully across the even theodolites:
furl]
ps_partkey: [71984], ps_suppkey: [6999], ps_avallqty: [6016], ps_supplycost: [1.01], ps_comment: [eodolites are blithely across the special requests. quickly regular excuses are furiously against the
slyly final accou]
ps_partkey: [74375], ps_suppkey: [6883], ps_avallqty: [864], ps_supplycost: [1.02], ps_comment: [lithely express asymptotes nag regular packages. special, ruthless instructions against the furiously
ruthless packages boost around the packages. slyly bold accounts use. furiously ironic pa]
ps_partkey: [76994], ps_suppkey: [9502], ps_avallqty: [9712], ps_supplycost: [1], ps_comment: [.] carefully ironic platelets cajole furiously among the furiously regular asymptotes. furiously express
asymptotes wake caref]
ps_partkey: [93653], ps_suppkey: [3654], ps_avallqty: [4473], ps_supplycost: [1.02], ps_comment: [ of the carefully final requests. bold deposits are slyly. instructions nod furiously instructions.
careful]
ps_partkey: [102497], ps_suppkey: [2498], ps_avallqty: [6491], ps_supplycost: [1], ps_comment: [fully final accounts. even accounts after the carefully final accounts haggle according to the blithely
special requests. carefully unusual]
ps_partkey: [122543], ps_suppkey: [5856], ps_avallqty: [5753], ps_supplycost: [1], ps_comment: [e the quickly ironic dependencies. slyly ironic accounts]
ps_partkey: [139711], ps_suppkey: [9712], ps_avallqty: [4286], ps_supplycost: [1.01], ps_comment: [ully unusual escapades sleep along the special instructions. final, bold ideas across the slyly ironic
ideas sleep dependenc]
ps_partkey: [155112], ps_suppkey: [5113], ps_avallqty: [7635], ps_supplycost: [1], ps_comment: [refully bold packages. special somas cajole according to the foxes. furiously even accou]
ps_partkey: [158893], ps_suppkey: [5639], ps_avallqty: [3751], ps_supplycost: [1], ps_comment: [lously unusual gifts maintain quickly according to the slyly pending deposits. quickly ]
ps_partkey: [193659], ps_suppkey: [6179], ps_avallqty: [6606], ps_supplycost: [1.01], ps_comment: [can haggle. quickly express packages are blithely. even requests against the silent accounts sleep
special packages. ironic ideas according to the furiously regular dolphins use quickly plate]
ps_partkey: [193981], ps_suppkey: [3982], ps_avallqty: [619], ps_supplycost: [1.01], ps_comment: [lc accounts after the unusual, regular instructions 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/tpch-dbgen/
heap files dir: bin/

runLength is:::0
number of records inserted::12
lnt: [31], string: [slate seashell steel medium moccasin], double: [931.03]
lnt: [1038], string: [orange floral olive ivory lace], double: [931.03]
lnt: [2029], string: [midnight brown dm violet almond], double: [931.02]
lnt: [3028], string: [puff slate tomato moccasin azure], double: [931.02]
lnt: [4027], string: [white ivory moccasin coral puff], double: [931.02]
lnt: [5026], string: [blanched blush pink light wheat], double: [931.02]
lnt: [6025], string: [purple medum light aquamarine dark], double: [931.02]
lnt: [7024], string: [forest rosy peach antique midnigh], double: [931.02]
lnt: [8023], string: [mint salmon moccasin blanched belge], double: [931.02]
lnt: [9022], string: [peru misty sandy dark drab], double: [931.02]
lnt: [10021], string: [blush steel green stenna snow], double: [931.02]
lnt: [11020], string: [plum khaki powder belge peru], double: [931.02]

query2 returned 12 records

*****

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

runLength is:::0
number of records inserted::10000
double: [9.24623e+07]

query3 returned 1 records

*****

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

query4
runLength is:::0
number of records inserted::10000
number of records inserted::800000
double: [4.40421e+08]
query4 returned 1 records
```

### 2. Output of Query 2, 3 and 4: [Snapshot of Output1.txt]

```
output1.txt (Work /media/gloria/Work1/UniversityOfFlorida/Study/Sem1/DB1/Projects/Project3/a3test) - gedit
Open  [F] Save

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

runLength is:::0
number of records inserted::12
lnt: [31], string: [slate seashell steel medium moccasin], double: [931.03]
lnt: [1038], string: [orange floral olive ivory lace], double: [931.03]
lnt: [2029], string: [midnight brown dm violet almond], double: [931.02]
lnt: [3028], string: [puff slate tomato moccasin azure], double: [931.02]
lnt: [4027], string: [white ivory moccasin coral puff], double: [931.02]
lnt: [5026], string: [blanched blush pink light wheat], double: [931.02]
lnt: [6025], string: [purple medum light aquamarine dark], double: [931.02]
lnt: [7024], string: [forest rosy peach antique midnigh], double: [931.02]
lnt: [8023], string: [mint salmon moccasin blanched belge], double: [931.02]
lnt: [9022], string: [peru misty sandy dark drab], double: [931.02]
lnt: [10021], string: [blush steel green stenna snow], double: [931.02]
lnt: [11020], string: [plum khaki powder belge peru], double: [931.02]

query2 returned 12 records

*****

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

runLength is:::0
number of records inserted::10000
double: [9.24623e+07]

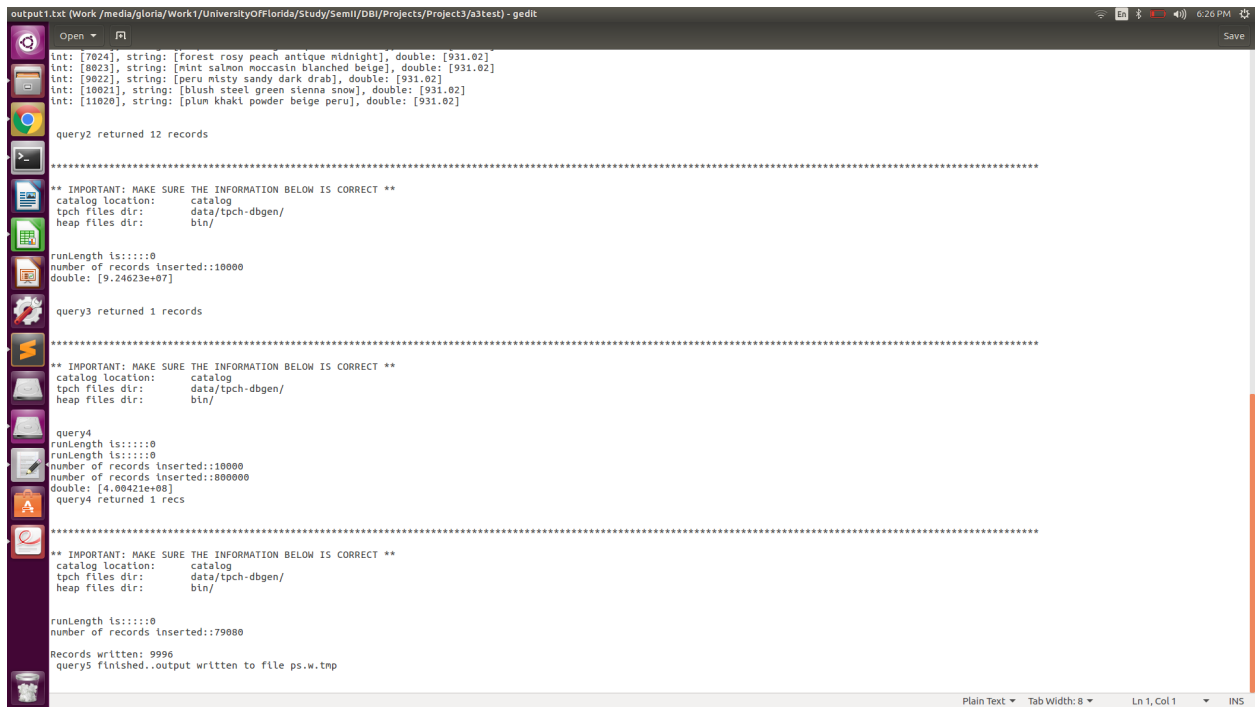
query3 returned 1 records

*****

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

query4
runLength is:::0
number of records inserted::10000
number of records inserted::800000
double: [4.40421e+08]
query4 returned 1 records
```

### 3. Output of Query 5: [Snapshot of Output1.txt]



```
output1.txt (Work /media/gloria/Work1/UniversityOfFlorida/Study/Sem1/DBI/Projects/Project3/a3test) - gedit
Open  [?]  Save

int: [7024], string: [forest rosy peach antique midnight], double: [931.02]
int: [8023], string: [mint salmon moccasin blanchet beige], double: [931.02]
int: [9022], string: [peru misty sandy dark drab], double: [931.02]
int: [10021], string: [blush steel green stenna 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 **
catalog location:      catalog
tpch files dir:       data/tpch-dbggen/
heap files dir:       bin/

runLength is:::0
number of records inserted::10000
double: [9.24623e+07]

query3 returned 1 records

*****

** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location:      catalog
tpch files dir:       data/tpch-dbggen/
heap files dir:       bin/

query4
runLength is:::0
runLength is:::0
number of records inserted::10000
number of records inserted::800000
double: [4.00421e+08]
query4 returned 1 recs

*****

** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location:      catalog
tpch files dir:       data/tpch-dbggen/
heap files dir:       bin/

runLength is:::0
number of records inserted::79000
Records written: 9996
query5 finished..output written to file ps.w.tmp

Plain Text  Tab Width: 8  Ln 1, Col 1  INS
```

### Gtest Results:

1. TEST(selectFile,test1)  
This test selects data from the nation bin file and filters for records whose region key is <3 and returns output(\_o). The count of records in the output pipe is matched in the assert.
2. TEST(selectPipe,test2)  
This test selects data from the nation bin file puts it in pipe then filters for records whose region key is >3 and returns output(\_o). The count of records in the output is matched in the assert.
3. TEST(sum,test3)  
This test selects data from the nation bin file then filters for records whose region key is >3. Post this it calculates the sum region keys of all the filtered records and returns output(\_out). If a single sum record is returned as expected, the test is pass.
4. TEST(duplicateRemoval,test4)  
This test case checks for unique values of n\_regionkey present in the table and adds it in the \_\_ps pipe. The count of the records is matched in the assert function.

Screenshot of Gtest.out

```
gloria@gloria-Inspiron-5593: /media/gloria/Work1/UniversityOfFlorida/Study/SemII/DBI/Projects/Project3/a3test$ ./gtest.out
** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location: catalog
tpch files dir: data/tpch-dbggen/
heap files dir: bin/

=====
Running 4 tests from 4 test cases.
Global test environment set-up.
-----
1 test from selectFile
selectFile.test1
run
runLength is:::0
number of records inserted::15
OK selectFile.test1 (3 ms)
-----
1 test from selectFile (3 ms total)

=====
1 test from selectPipe
selectPipe.test2
run
runLength is:::0
number of records inserted::5
OK selectPipe.test2 (2 ms)
-----
1 test from selectPipe (2 ms total)

=====
1 test from sum
sum.test3
run
runLength is:::0
number of records inserted::5
Int: [28]
OK sum.test3 (2 ms)
-----
1 test from sum (2 ms total)

=====
1 test from duplicateRemoval
duplicateRemoval.test4
run
runLength is:::0
number of records inserted::5
OK duplicateRemoval.test4 (17 ms)
-----
1 test from duplicateRemoval (17 ms total)

=====
Global test environment tear-down
=====
4 tests from 4 test cases ran. (24 ms total)
PASSED 4 tests.
gloria@gloria-Inspiron-5593: /media/gloria/Work1/UniversityOfFlorida/Study/SemII/DBI/Projects/Project3/a3test$
```

## Issues in Code:

Could not run the q6 function given in test.cc as the OrderMaker required for groupBy was not defined. Defined the OrderMaker adding the below mentioned lines of code and executed it successfully.

```
OrderMaker grp_order;
grp_order.numAtts=1;
int n = join_sch.GetNumAtts();
Attribute *myAtts=join_sch.GetAtts();
for(int i=0;i<n;i++)
{
    if(i==3)
    {
        grp_order.whichAtts[0]=i;
        grp_order.whichTypes[0]=Int;
    }
}
```

gloria@gloria-inspiron-5593: /media/loria/Work1/UniversityOfFlorida/Study/SemI/DBI/Projects/Project3/a3test  
gloria@gloria-inspiron-5593: /media/loria/Work1/UniversityOfFlorida/Study/SemI/DBI/Projects/Project3/a3test\$ ./test.out 6

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

catalog location: catalog  
tpch files dir: data/tpch-dbggen/  
heap files dir: bin/

query6

runLength ts:::0  
number of records inserted::10000  
number of records inserted::800000  
double: [1.68363e+07]  
double: [1.65363e+07]  
double: [1.59263e+07]  
double: [1.6502e+07]  
double: [1.65789e+07]  
double: [1.52194e+07]  
double: [1.61179e+07]  
double: [1.57786e+07]  
double: [1.65973e+07]  
double: [1.62349e+07]  
double: [1.56944e+07]  
double: [1.75655e+07]  
double: [1.51488e+07]  
double: [1.45051e+07]  
double: [1.50595e+07]  
double: [1.4906e+07]  
double: [1.62842e+07]  
double: [1.68236e+07]  
double: [1.63086e+07]  
double: [1.59509e+07]  
double: [1.64843e+07]  
double: [1.59514e+07]  
double: [1.60954e+07]  
double: [1.5516e+07]  
double: [1.57613e+07]  
query6 returned sum for 25 groups (expected 25 groups)

gloria@gloria-inspiron-5593: /media/loria/Work1/UniversityOfFlorida/Study/SemI/DBI/Projects/Project3/a3test\$