Assignment 4.1: Statistical Estimation

Himavanth Boddu 32451847
Akib Maredia 38856489

Steps to run

- Unzip the compressed file using akibMaredia_himavanthBoddu_p41.zip
- cd akibMaredia_himavanthBoddu_p41
- make a2test.out for generating dependencies and run ./a2test.out.
- `make test.out` to compile test.out("./test.out [1-11]" to run the compiled binary where "[1-11]" denotes the query number you want to execute)
- `./runTestCases.sh` to test the output
- `make gtest` to compile Gtest.cc
- `./gtest` to run test cases

Implementation

We have defined a structure named statisticSchema with integer values storing number of tuples and number of relations and a map for storing attributes.
We are using relationMap map structure for quickly locating attributes and relations.
We also implemented a constructor Statistics (), a copy constructor Statistics (Statistics &copyMe) which performs deep copy and populates newStatisticSchema object which is inserted into relationMap, and a destructor ~Statistics () where we clear the relationMap

Container Operations:

- int AddRel (char *relName, int numTuples) : This function adds new relation specified by name and number of tuples. If relation name is not in relation map then we add a new statisticSchema else update the numTuples. It returns zero if relation name is passed as null else returns one.
- void AddAtt (char* relName, char *attName, int numDistincts): This function adds attribute to one of relations by specifying relation name, what relation the attribute is attached to, and the number of distinct values that the relation has for that particular attribute. If new attribute name is not in attributes map then add the new attribute else update number of matches.
- void CopyRel (char *oldName, char *newName): This function creates the copy of the relation and saves it with new name. It can also write into text file and read back from it. If the object has to read itself from the file which is not present we get empty output.
- void Read (char *fromWhere) and void Write (char *fromWhere) for reading and writing input and output respectively to statistics.txt and output41.txt.

What-If Operations

- void Apply(struct AndList *parseTree, char **relNames, int numToJoin): This function simulates join of all relations present in relNames variable using predicates from parseTree variable. We do not actually implement join as it is done in RelOps class. We populate new statisticSchema joinedRel and add it to relationMap until records get exhausted in the case if join is possible.

- int Estimate(struct AndList *parseTree, char **relNames, int numToJoin); This operation is exactly like Apply, except that it does not actually change the state of the Statistics object. Instead, it computes the number of tuples that would result from a join over the relations in relNames, and returns this to the caller. Here we save the state as most left then the value is zero else one. We are using logic to truncate the output obtained in double value as it has to be a whole number.

How to interpret statistics.txt

In the write method we have defined the pattern for writing the output in the following format
In the first line we have the relation mentioned and the number as output is the result we get from estimate method in the following format
Relation <relation_name> < Estimated tuple count after performing join >

In the following lines we mention the attributes and the distinct count of attribute records
Attribute <attribute_name > <distinct count of attribute>

Bugs in test.cc

i. Spelling mistake in q3 line s.AddAtt(relName[0], "s_nationey", 25); should be s.AddAtt(relName[0], "s_nationkey", 25);
ii. Wrong relations given as parameters in q4, corresponding changes have been done in the test.cc.
iii. Attribute was not added for relation “order” as “o_orderdate” in q5. Added the line s.AddAtt(relName[1], “o_orderdate”, 99996);
iv. Attribute was not added for the relation “lineitem” as “l_receiptdate” in q7. Added the line s.AddAtt(relName[1], “l_receiptdate”, 198455);
v. Attribute was not added for relation “order” as “o_orderdate” in q10. Added line s.AddAtt(relName[1], “o_orderdate”, 99996);
vi. There should be “yyparse();” at 2 places in q10.
vii. Spelling mistake in q11 “p_conatiner” should be “p_container”

Results:

- ## Gtest

```
himavanthboddu@himavanths-MacBook-Air ~/Documents/akibMaredia_himavanthBoddu_p41/42  master ● ./gtest
[==========] Running 2 tests from 2 test suites.
[----------] Global test environment set-up.
[----------] 1 test from AddRelFail
[ RUN      ] AddRelFail.Run
[       OK ] AddRelFail.Run (0 ms)
[----------] 1 test from AddRelFail (0 ms total)

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

[----------] Global test environment tear-down
[==========] 2 tests from 2 test suites ran. (0 ms total)
[  PASSED  ] 2 tests.
```

Output

```
Relation lineitem 857316
Attribute l_discount 11
Attribute l_returnflag 3
Attribute l_shipmode 7
********************************************************************************************************************************************************
Relation customer|orders|nation 1500000
Attribute c_custkey 150000
Attribute c_nationkey 25
Attribute n_nationkey 25
Attribute o_custkey 150000
********************************************************************************************************************************************************
Relation lineitem|customer|orders 400081
Attribute c_custkey 150000
Attribute c_mktsegment 5
Attribute l_orderkey 1500000
Attribute o_custkey 150000
Attribute o_orderdate 99996
Attribute o_orderkey 1500000
********************************************************************************************************************************************************
Relation lineitem|customer|orders|nation 2000405
Attribute c_custkey 150000
Attribute c_nationkey 25
Attribute l_orderkey 1500000
Attribute n_nationkey 25
Attribute o_custkey 150000
Attribute o_orderdate 99996
Attribute o_orderkey 1500000
********************************************************************************************************************************************************
Relation lineitem|part 21432
Attribute l_partkey 200000
Attribute l_shipinstruct 4
Attribute l_shipmode 7
Attribute p_container 40
Attribute p_partkey 200000
********************************************************************************************************************************************************
```