

Database System Implementation

Project 4 -2 – Query Compiler and Optimizer

Developed by –

Ghodkari Chowdary, Raghunatha Rao - UFID: 6218-1051
Mullapudi, Aseesh - UFID: 9175-1971

Aim of this project is to extend our Database System with a query compiler and optimizer. Previous we have computed statistical estimation for the relations, now we extend the database to use these estimations in generating optimized query plans for execution of the SQL queries. We have completed the code in the main.cc file, which takes the input SQL query and processes it and generates the optimized query plan for execution.

Implementation Details:

Explaining the Output: For TestCase1

SELECT n.n_nationkey FROM nation AS n WHERE (n.n_name = 'UNITED STATES')

For each operation, its corresponding PipeID input and output are printed. After the operation corresponding the supporting operations are printed. For e.g. Output schema for SELECT, PIPE and PROJECT operations. CNF of predicate for SELECT and JOIN operations. Aggregate function for both SUM and GROUP BY. Ordermaker is printed for Group BY operation.

| | |
|----------------------------|---|
| Number of selects: 1 | (Number of Select Operations in the Predicate) |
| Number of joins: 0 | (Number of Join Operations in the Predicate) |
| Att n.n_nationkey: INT | (A query tree is built over the SQL query) |
| PRINTING TREE IN ORDER: | (Printing the Query Tree in order) |
| ***** | |
| SELECT FILE operation | (first is the SELECT operation over nation table) |
| Input Pipe 0 | |
| Output Pipe 1 | |
| Output Schema: | |
| Att n.n_nationkey: INT | |
| Att n.n_name: STRING | |
| Att n.n_regionkey: INT | |
| Att n.n_comment: STRING | |
| ***** | |
| SELECT PIPE operation | (Pipe where predicate n_name = 'UNITED STATES' is applied) |
| Input Pipe 1 | |
| Output Pipe 2 | |
| Output Schema: | |
| Att n.n_nationkey: INT | |
| Att n.n_name: STRING | |
| Att n.n_regionkey: INT | |
| Att n.n_comment: STRING | |
| SELECTION CNF : | |
| (n.n_name = UNITED STATES) | |
| ***** | |
| PROJECT operation | (Final Project operation where the required column is filtered) |
| Input Pipe 2 | |
| Output Pipe 3 | |
| Output Schema: | |
| Att n.n_nationkey: INT | |

Main.cc Code Structure:

- Initially the input sql query is processed into data structures using `yyparse()`, it builds the required data structures for further processing
- We read and initialize the list of relations in the query
- ***getPredicateInfo(joinsVector, selectsVector, joinDepSelectsVector, *statsFileObj)***
 - Function reads the existing data structure, especially the `AndList` where the predicate of the query is stored. It separates the different kinds of CNF in the predicates into Join, Select and Select dependent on Join operations.
- Map is built for storing the Query tree for each relation involved in the query
map<string, QueryTreeNode*> leaves
- Now, firstly SELECT operations are used for estimation and the new nodes into the query tree are created as top nodes to query tree over the existing query tree nodes for corresponding relations
- If joins exist in predicate, we evaluate optimized way to compute the Join operations
- ***optimizeOrderofJoins(joinsVector, statsFileObj)***
 - Function implements the Left deep Join optimization; all possible join combinations are evaluated, and a minimum cost value is used and depending on it an optimized way to execute joins is achieved
- Join operations processed in the optimized order and the new nodes into the query tree are created as top nodes to query tree over the existing query tree nodes for corresponding relations
- Select operations dependent on the Joins are processed accordingly, Later, if Aggregated Functions exist in the query they processed based on operations (Sum, Distinct, Group By).
- Later the Distinct, Project operations are added to the query tree accordingly if exist
- ***void QueryTreeNode::printInOrder()***
 - Query plan is generated using a Post order traversal over the query tree

(GTests) gtest.cc:

This file implements google tests for 4 testcases using the functions that count the number of relations, number of joins, number of select predicates and number of selects depend on joins. The test cases also extended over the function which implement the optimization logic over the joins.

Instructions to Run the Code:

To run the overall project, follow the below steps:

Compile and run main.cc:

Ensure there is Catalog file, Statistics file in the root folder as well

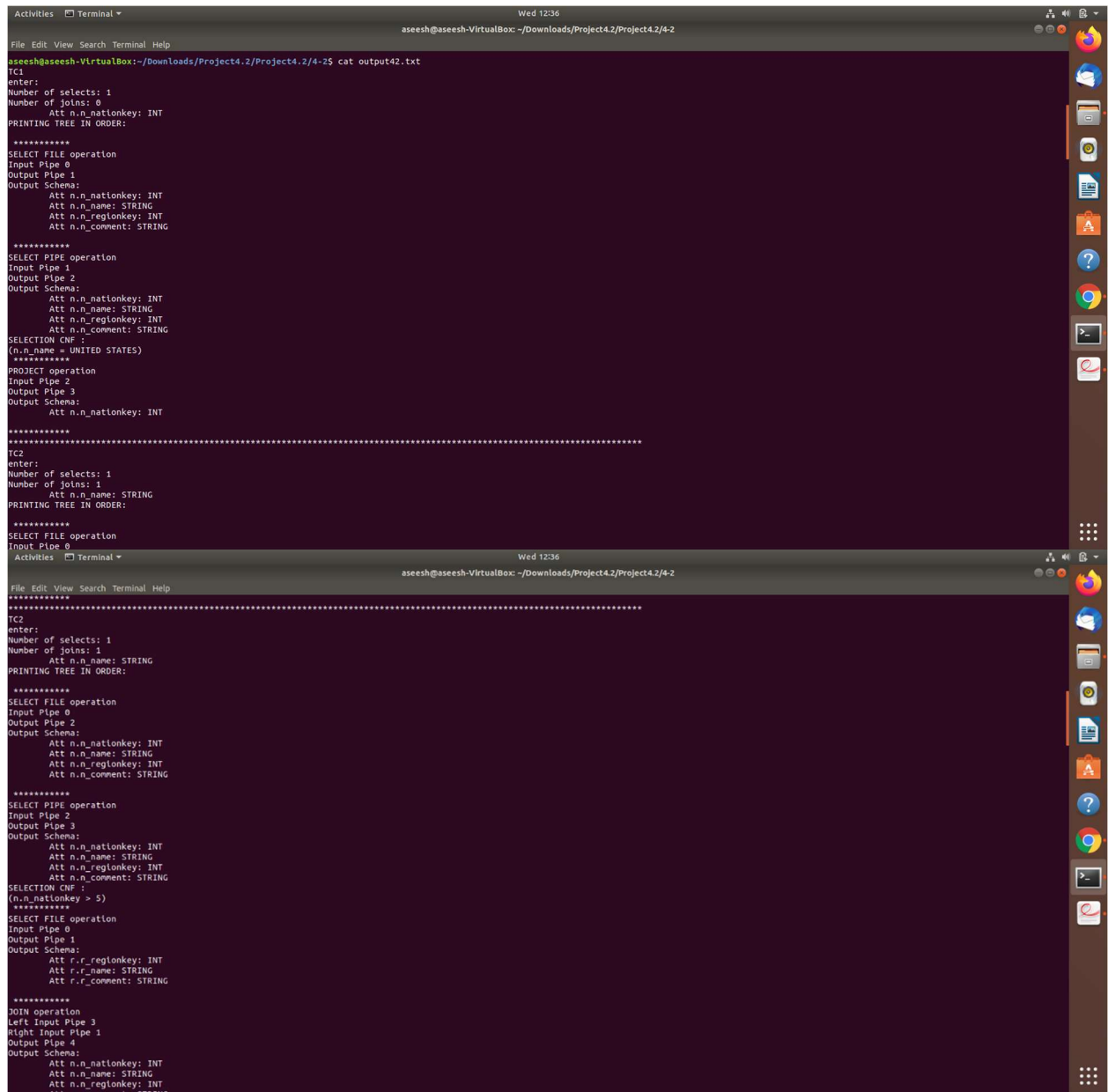
1. `make`
2. `./a42.out`
3. Enter the SQL query on the console and press CTRL D

To run the test cases

1. make
2. ./runTestCases42.sh

File output42.txt is generated

The screenshot of the output42.txt is below



```
File Edit View Search Terminal Help
aseesh@aseesh-VirtualBox: ~/Downloads/Project4.2/Project4.2/4-2$ cat output42.txt
TC1
enter:
Number of selects: 1
Number of joins: 0
Att n.n_nationkey: INT
PRINTING TREE IN ORDER:
*****
SELECT FILE operation
Input Pipe 0
Output Pipe 1
Output Schema:
Att n.n_nationkey: INT
Att n.n_name: STRING
Att n.n_regionkey: INT
Att n.n_comment: STRING
*****
SELECT PIPE operation
Input Pipe 1
Output Pipe 2
Output Schema:
Att n.n_nationkey: INT
Att n.n_name: STRING
Att n.n_regionkey: INT
Att n.n_comment: STRING
SELECTION CNF :
(n.n_name = UNITED STATES)
*****
PROJECT operation
Input Pipe 2
Output Pipe 3
Output Schema:
Att n.n_nationkey: INT
*****
TC2
enter:
Number of selects: 1
Number of joins: 1
Att n.n_name: STRING
PRINTING TREE IN ORDER:
*****
SELECT FILE operation
Input Pipe 0
Output Pipe 2
Output Schema:
Att n.n_nationkey: INT
Att n.n_name: STRING
Att n.n_regionkey: INT
Att n.n_comment: STRING
*****
SELECT PIPE operation
Input Pipe 2
Output Pipe 3
Output Schema:
Att n.n_nationkey: INT
Att n.n_name: STRING
Att n.n_regionkey: INT
Att n.n_comment: STRING
SELECTION CNF :
(n.n_nationkey > 5)
*****
SELECT FILE operation
Input Pipe 0
Output Pipe 1
Output Schema:
Att r.r_regionkey: INT
Att r.r_name: STRING
Att r.r_comment: STRING
*****
JOIN operation
Left Input Pipe 3
Right Input Pipe 1
Output Pipe 4
Output Schema:
Att n.n_nationkey: INT
Att n.n_name: STRING
Att n.n_regionkey: INT
Att n.n_comment: STRING
```

```
Activities Terminal
Wed 12:37
aseesh@aseesh-VirtualBox: ~/Downloads/Project4.2/Project4.2/4-2

File Edit View Search Terminal Help

*****
JOIN operation
Left Input Pipe 3
Right Input Pipe 1
Output Pipe 4
Output Schema:
    Att n.n_nationkey: INT
    Att n.n_name: STRING
    Att n.n_regionkey: INT
    Att n.n_comment: STRING
    Att r.r_regionkey: INT
    Att r.r_name: STRING
    Att r.r_comment: STRING

CNF:
(n.n_regionkey = r.r_regionkey)
*****
PROJECT operation
Input Pipe 4
Output Pipe 3
Output Schema:
    Att n.n_name: STRING
*****
TC3
enter:
Number of selects: 1
Number of joins: 1
PRINTING TREE IN ORDER:
*****
SELECT FILE operation
Input Pipe 0
Output Pipe 2
Output Schema:
    Att n.n_nationkey: INT
    Att n.n_name: STRING
    Att n.n_regionkey: INT
    Att n.n_comment: STRING
*****
SELECT PIPE operation
Input Pipe 2
Output Pipe 3
Output Schema:
    Att n.n_nationkey: INT
*****
TC3
enter:
Number of selects: 1
Number of joins: 1
PRINTING TREE IN ORDER:
*****
SELECT FILE operation
Input Pipe 0
Output Pipe 2
Output Schema:
    Att n.n_nationkey: INT
    Att n.n_name: STRING
    Att n.n_regionkey: INT
    Att n.n_comment: STRING
*****
SELECT PIPE operation
Input Pipe 2
Output Pipe 3
Output Schema:
    Att n.n_nationkey: INT
    Att n.n_name: STRING
    Att n.n_regionkey: INT
    Att n.n_comment: STRING
*****
SELECTION CNF 1
(n.n_name = UNITED STATES)
*****
SELECT FILE operation
Input Pipe 0
Output Pipe 1
Output Schema:
    Att r.r_regionkey: INT
    Att r.r_name: STRING
    Att r.r_comment: STRING
*****
JOIN operation
Left Input Pipe 3
Right Input Pipe 1
Output Pipe 4
Output Schema:
    Att n.n_nationkey: INT
    Att n.n_name: STRING
    Att n.n_regionkey: INT
    Att n.n_comment: STRING
```

```
Activities Terminal
Wed 12:37
aseesh@aseesh-VirtualBox: ~/Downloads/Project4.2/Project4.2/4-2

File Edit View Search Terminal Help
*****
JOIN operation
Left Input Pipe 3
Right Input Pipe 1
Output Pipe 4
Output Schema:
    Att n.n_nationkey: INT
    Att n.n_name: STRING
    Att n.n_regionkey: INT
    Att n.n_comment: STRING
    Att r.r_regionkey: INT
    Att r.r_name: STRING
    Att r.r_comment: STRING

CNF:
(n.n_regionkey = r.r_regionkey)
*****
SUM operation
Left Input Pipe 4
Output Pipe 3
Output Schema:
    Att n.n_nationkey: INT
    Att n.n_name: STRING
    Att n.n_regionkey: INT
    Att n.n_comment: STRING
    Att r.r_regionkey: INT
    Att r.r_name: STRING
    Att r.r_comment: STRING

FUNCTION:
NumOps = 1
PushInt
returnsInt
*****
TC4
enter:
Number of selects: 1
Number of joins: 1
GROUPING ON n.n_regionkey
PRINTING TREE IN ORDER:
*****
SELECT FILE operation
Input Pipe 2
Output Pipe 2
Output Schema:
    Att n.n_nationkey: INT
    Att n.n_name: STRING

*****
SELECT PIPE operation
Input Pipe 2
Output Pipe 3
Output Schema:
    Att n.n_nationkey: INT
    Att n.n_name: STRING
    Att n.n_regionkey: INT
    Att n.n_comment: STRING

SELECTION CNF:
(n.n_name = UNITED STATES)
*****
SELECT FILE operation
Input Pipe 0
Output Pipe 1
Output Schema:
    Att r.r_regionkey: INT
    Att r.r_name: STRING
    Att r.r_comment: STRING

*****
JOIN operation
Left Input Pipe 3
Right Input Pipe 1
Output Pipe 4
Output Schema:
    Att n.n_nationkey: INT
    Att n.n_name: STRING
    Att n.n_regionkey: INT
    Att n.n_comment: STRING
```

```
Activities Terminal
Wed 12:38
aseesh@aseesh-VirtualBox: ~/Downloads/Project4.2/Project4.2/4-2

File Edit View Search Terminal Help

*****
JOIN operation
Left Input Pipe 3
Right Input Pipe 1
Output Pipe 4
Output Schema:
    Att n.n.nationkey: INT
    Att n.n.name: STRING
    Att n.n.regionkey: INT
    Att n.n.comment: STRING
    Att r.r.regionkey: INT
    Att r.r.name: STRING
    Att r.r.comment: STRING
CNF:
(n.n.regionkey = r.r.regionkey)
*****
GROUP BY operation
Left Input Pipe 4
Output Pipe 5
Output Schema:
    Att n.n.nationkey: INT
    Att n.n.name: STRING
    Att n.n.regionkey: INT
    Att n.n.comment: STRING
    Att r.r.regionkey: INT
    Att r.r.name: STRING
    Att r.r.comment: STRING
GROUPING ON:
NUMatts = 1
0: 2 Int
FUNCTION:
NUMops = 1
pushInt
returnsInt
*****
enter:
Number of selects: 1
Number of joins: 2
GROUPING ON r.r.regionkey
PRINTING TREE IN ORDER:
*****
SELECT FILE operation
*****
Activities Terminal
Wed 12:38
aseesh@aseesh-VirtualBox: ~/Downloads/Project4.2/Project4.2/4-2

File Edit View Search Terminal Help

*****
TCS
enter:
Number of selects: 1
Number of joins: 2
GROUPING ON r.r.regionkey
PRINTING TREE IN ORDER:
*****
SELECT FILE operation
Input Pipe 0
Output Pipe 3
Output Schema:
    Att n.n.nationkey: INT
    Att n.n.name: STRING
    Att n.n.regionkey: INT
    Att n.n.comment: STRING
*****
SELECT PIPE operation
Input Pipe 3
Output Pipe 4
Output Schema:
    Att n.n.nationkey: INT
    Att n.n.name: STRING
    Att n.n.regionkey: INT
    Att n.n.comment: STRING
SELECTION CNF :
(n.n.nationkey > 10)
*****
SELECT FILE operation
Input Pipe 0
Output Pipe 2
Output Schema:
    Att r.r.regionkey: INT
    Att r.r.name: STRING
    Att r.r.comment: STRING
*****
JOIN operation
Left Input Pipe 4
Right Input Pipe 2
Output Pipe 5
Output Schema:
    Att n.n.nationkey: INT
    Att n.n.name: STRING
    Att n.n.regionkey: INT
    Att n.n.comment: STRING
```

```
Activities Terminal
Wed 12:39
aseesh@aseesh-VirtualBox: ~/Downloads/Project4.2/Project4.2/4-2

File Edit View Search Terminal Help

*****
JOIN operation
Left Input Pipe 4
Right Input Pipe 2
Output Pipe 5
Output Schema:
Att n.n_nationkey: INT
Att n.n_name: STRING
Att n.n_regionkey: INT
Att n.n_comment: STRING
Att r.r_regionkey: INT
Att r.r_name: STRING
Att r.r_comment: STRING

CNF:
(n.n_regionkey = r.r_regionkey)
*****
SELECT FILE operation
Input Pipe 0
Output Pipe 1
Output Schema:
Att c.c_custkey: INT
Att c.c_name: STRING
Att c.c_address: STRING
Att c.c_nationkey: INT
Att c.c_phone: STRING
Att c.c_acctbal: DOUBLE
Att c.c_aktsegment: STRING
Att c.c_comment: STRING

*****
JOIN operation
Left Input Pipe 5
Right Input Pipe 1
Output Pipe 6
Output Schema:
Att n.n_nationkey: INT
Att n.n_name: STRING
Att n.n_regionkey: INT
Att n.n_comment: STRING
Att r.r_regionkey: INT
Att r.r_name: STRING
Att r.r_comment: STRING
Att c.c_custkey: INT
Att c.c_name: STRING
Att c.c_address: STRING
Att c.c_nationkey: INT
Att c.c_phone: STRING

Activities Terminal
Wed 12:39
aseesh@aseesh-VirtualBox: ~/Downloads/Project4.2/Project4.2/4-2

File Edit View Search Terminal Help

*****
JOIN operation
Left Input Pipe 5
Right Input Pipe 1
Output Pipe 6
Output Schema:
Att n.n_nationkey: INT
Att n.n_name: STRING
Att n.n_regionkey: INT
Att n.n_comment: STRING
Att r.r_regionkey: INT
Att r.r_name: STRING
Att r.r_comment: STRING
Att c.c_custkey: INT
Att c.c_name: STRING
Att c.c_address: STRING
Att c.c_nationkey: INT
Att c.c_phone: STRING
Att c.c_acctbal: DOUBLE
Att c.c_aktsegment: STRING
Att c.c_comment: STRING

CNF:
(n.n_nationkey = c.c_nationkey)
*****
DISTINCT operation
Left Input Pipe 6
Output Pipe 7
Output Schema:
Att n.n_nationkey: INT
Att n.n_name: STRING
Att n.n_regionkey: INT
Att n.n_comment: STRING
Att r.r_regionkey: INT
Att r.r_name: STRING
Att r.r_comment: STRING
Att c.c_custkey: INT
Att c.c_name: STRING
Att c.c_address: STRING
Att c.c_nationkey: INT
Att c.c_phone: STRING
Att c.c_acctbal: DOUBLE
Att c.c_aktsegment: STRING
Att c.c_comment: STRING

*****
GROUP BY operation
Left Input Pipe 7
Output Pipe 8
```



```
Activities Terminal
Wed 12:39
aseesh@aseesh-VirtualBox: ~/Downloads/Project4.2/Project4.2/4-2

File Edit View Search Terminal Help
Att n.n.nationkey: INT
Att n.n.name: STRING
Att n.n.regionkey: INT
Att n.n.comment: STRING
Att r.r.regionkey: INT
Att r.r.name: STRING
Att r.r.comment: STRING
Att c.c.custkey: INT
Att c.c.name: STRING
Att c.c.address: STRING
Att c.c.nationkey: INT
Att c.c.phone: STRING
Att c.c.acctbal: DOUBLE
Att c.c.mktsegment: STRING
Att c.c.comment: STRING
*****
GROUP BY operation
Left Input Pipe 7
Output Pipe 8
Output Schema:
Att n.n.nationkey: INT
Att n.n.name: STRING
Att n.n.regionkey: INT
Att n.n.comment: STRING
Att r.r.regionkey: INT
Att r.r.name: STRING
Att r.r.comment: STRING
Att c.c.custkey: INT
Att c.c.name: STRING
Att c.c.address: STRING
Att c.c.nationkey: INT
Att c.c.phone: STRING
Att c.c.acctbal: DOUBLE
Att c.c.mktsegment: STRING
Att c.c.comment: STRING
*****
GROUPING ON:
NUMatts = 1
0: 4 Int
*****
FUNCTION:
NumOps = 3
PushInt
PushInt
IntPlus
returnsInt
*****
aseesh@aseesh-VirtualBox:~/Downloads/Project4.2/Project4.2/4-2
```

Instructions to run gtests:

Compile and run code gtest.cc:

1. make gtest
2. ./gtest

The screenshot of ./gtest is below

```
Activities Terminal
Wed 12:39
aseesh@aseesh-VirtualBox: ~/Downloads/Project4.2/Project4.2/4-2

File Edit View Search Terminal Help
aseesh@aseesh-VirtualBox:~/Downloads/Project4.2/Project4.2/4-2$ ./gtest
[*****] Running 4 tests from 4 test cases.
[-----] Global test environment set-up.
[-----] 1 test from TestCase1
[ RUN ] TestCase1.SubTest1
[ OK ] TestCase1.SubTest1 (0 ms)
[-----] 1 test from TestCase1 (0 ms total)
[-----] 1 test from TestCase2
[ RUN ] TestCase2.SubTest2
[ OK ] TestCase2.SubTest2 (0 ms)
[-----] 1 test from TestCase2 (0 ms total)
[-----] 1 test from TestCase3
[ RUN ] TestCase3.SubTest3
[ OK ] TestCase3.SubTest3 (0 ms)
[-----] 1 test from TestCase3 (0 ms total)
[-----] 1 test from TestCase4
[ RUN ] TestCase4.SubTest4
[ OK ] TestCase4.SubTest4 (0 ms)
[-----] 1 test from TestCase4 (0 ms total)
[-----] Global test environment tear-down
[*****] 4 tests from 4 test cases ran. (0 ms total)
[ PASSED ] 4 tests.
aseesh@aseesh-VirtualBox:~/Downloads/Project4.2/Project4.2/4-2$
```