

COP 6726 – Database System Implementation

Project 4_2: Query Compilation and Optimization

Group Members:

Sanjay Reddy Banda, UF ID: 5878-2239

Suprith Reddy Gurudu, UF ID: 9961-2134

Compilation and Execution:

To compile the code, run the following command:

```
>> make
```

To execute the main.cc code, change the directory to the specific folder (a4-1test) and run the following command:

```
>> ./a42.out
```

```
>> <Enter Query>
```

And press CTRL + D

Or

```
>> ./a42.out < testcase.sql
```

To compile the gTest (gtest.cc) code, run the following command:

```
>> make gtest.out
```

To execute the gTest (gtest.cc) code, run the following command:

```
>> ./gtest.out
```

Code Explanation (modified methods):

Filename: main.cc

Classname: QueryNode

Methods:

QueryNode():

Default constructor for QueryNode class.

QueryNode(NodeType type):

Parameterized constructor for QueryNode class with NodeType as argument.

virtual void Print():

Virtual function to print Node to stdout.

Classname: JoinNode <- QueryNode

Methods:

JoinNode():

Default constructor for JoinNode class.

void Print():

Function to print Join related attributes to stdout.

Classname: ProjectNode <- QueryNode

Methods:

ProjectNode ():

Default constructor for *ProjectNode* class.

void Print():

Function to print Project related attributes to stdout.

Classname: SelectFileNode <- QueryNode

Methods:

SelectFileNode ():

Default constructor for SelectFileNode class.

void Print():

Function to print SelectFile related attributes to stdout.

Classname: SelectPipeNode <- QueryNode

Methods:

SelectPipeNode ():

Default constructor for SelectPipeNode class.

void Print():

Function to print SelectPipe related attributes to stdout.

Classname: SumNode <- QueryNode

Methods:

SumNode ():

Default constructor for SumNode class.

void Print():

Function to print SumNode related attributes to stdout.

Classname: DistinctNode <- QueryNode

Methods:

DistinctNode ():

Default constructor for DistinctNode class.

void Print():

Function to print Distinct related attributes to stdout.

Classname: GroupByNode <- QueryNode

Methods:

GroupByNode ():

Default constructor for GroupByNode class.

void Print():

Function to print GroupBy related attributes to stdout.

Classname: WriteOutNode <- QueryNode

Methods:

WriteOutNode ():

Default constructor for WriteOutNode class.

void Print():

Function to print WriteOut related attributes to stdout.

void initSchemaMap(SchemaMap &map):

Creates the schema object for all the tables and inserts the objects into the map.

void initStatistics (Statistics &s):

Initializes the Statistics objects by adding all the relations and appropriate attributes.

*void CopyTablesNamesAndAliases (TableList *tableList, Statistics &s, vector<char *> &tableNames, AliaseMap &map):*

This function copies the table names and aliases.

*void CopyNameList(NameList *nameList, vector<string> &names):*

This function copies the name list.

Filename: gtest.cc

TEST(EXECUTIONPLAN, TC1) -

Google test for validating the test case for "SELECT n.n_nationkey FROM nation AS n WHERE (n.n_name = 'UNITED STATES')"

scenario. It verifies by following execution order "Project<-Select_File".

TEST(EXECUTIONPLAN, TC2) -

Google test for validating the test case for "SELECT n.n_name FROM nation AS n, region AS r WHERE (n.n_regionkey = r.r_regionkey) AND (n.n_nationkey > 5)" scenario. It verifies by following execution order "Project<-Select_File<-Select_File<-Join".

TEST(EXECUTIONPLAN, TC3) -

Google test for validating the test case for "SELECT SUM (n.n_nationkey) FROM nation AS n, region AS r WHERE (n.n_regionkey = r.r_regionkey) AND (n.n_name = 'UNITED STATES')"

scenario. It verifies by following execution order "Sum<-Select_File<-Select_File<-Join".

TEST(EXECUTIONPLAN, TC4) -

Google test for validating the test case for "SELECT SUM (n.n_regionkey) FROM nation AS n, region AS r WHERE (n.n_regionkey = r.r_regionkey) AND (n.n_name = 'UNITED STATES') GROUP BY n.n_regionkey"

scenario. It verifies by following execution order "Group_By<-Select_File<-Select_File<-Join".

TEST(EXECUTIONPLAN, TC5) -

Google test for validating the test case for "SELECT SUM DISTINCT (n.n_nationkey + r.r_regionkey) FROM nation AS n, region AS r, customer AS c WHERE (n.n_regionkey = r.r_regionkey) AND (n.n_nationkey = c.c_nationkey) AND (n.n_nationkey > 10) GROUP BY r.r_regionkey"

scenario. It verifies by following execution order "Group_By<-Duplicate_Elimination<-Select_File<-Select_File<-Join<-Select_File<-Join".

Results for the Test Cases from output42.txt:

Test Case 1:

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

```

a4-2test > E output42.txt
1 TC1
2 *****
3 Select File Operation
4 Output Pipe ID 1
5 Output Schema:
6     n.n_nationkey : Int
7     n.n_name : String
8     n.n_regionkey : Int
9     n.n_comment : String
10 Select CNF:
11 ( Att 1 from left record = Att 0 from literal record (String))
12 *****
13 *****
14 Project Operation
15 Input Pipe ID : 1
16 Output Pipe ID 2
17 Number Attrs Input : 4
18 Number Attrs Output : 1
19 Attrs To Keep :
20 0
21 *****
22 *****

```

Test Case 2:

Input- "SELECT n.n_name FROM nation AS n, region AS r WHERE (n.n_regionkey = r.r_regionkey) AND (n.n_nationkey > 5)"

```

a4-2test > E output42.txt
22 *****
23 TC2
24 *****
25 Select File Operation
26 Output Pipe ID 1
27 Output Schema:
28     n.n_nationkey : Int
29     n.n_name : String
30     n.n_regionkey : Int
31     n.n_comment : String
32 Select CNF:
33 ( Att 0 from left record > Att 0 from literal record (Int))
34 *****
35 *****
36 Select File Operation
37 Output Pipe ID 3
38 Output Schema:
39     r.r_regionkey : Int
40     r.r_name : String
41     r.r_comment : String
42 Select CNF:
43
44 *****
45 *****
46 Join Operation
47 Input Pipe 1 ID : 1
48 Input Pipe 2 ID : 3
49 Output Pipe ID : 2
50 Output Schema :
51     n.n_nationkey : Int
52     n.n_name : String
53     n.n_regionkey : Int
54     n.n_comment : String
55     r.r_regionkey : Int
56     r.r_name : String
57     r.r_comment : String
58 Join CNF :
59 ( Att 2 from left record = Att 0 from right record (Int)) AND
60 ( Att 0 from left record > Att 0 from literal record (Int))
61 *****
62 *****
63 Project Operation
64 Input Pipe ID : 2
65 Output Pipe ID 4
66 Number Attrs Input : 7
67 Number Attrs Output : 1
68 Attrs To Keep :
69 0
70 *****

```

Test Case 3:

Input- "SELECT SUM (n.n_nationkey) FROM nation AS n, region AS r WHERE (n.n_regionkey = r.r_regionkey) AND (n.n_name = 'UNITED STATES')"

```
a4-2test > ≡ output42.txt
72 | TC3
73 | *****
74 | Select File Operation
75 | Output Pipe ID 1
76 | Output Schema:
77 |     n.n_nationkey : Int
78 |     n.n_name : String
79 |     n.n_regionkey : Int
80 |     n.n_comment : String
81 | Select CNF:
82 | ( Att 1 from left record = Att 0 from literal record (String))
83 | *****
84 | *****
85 | Select File Operation
86 | Output Pipe ID 3
87 | Output Schema:
88 |     r.r_regionkey : Int
89 |     r.r_name : String
90 |     r.r_comment : String
91 | Select CNF:
92 |
93 | *****
94 | *****
95 | Join Operation
96 | Input Pipe 1 ID : 1
97 | Input Pipe 2 ID : 3
98 | Output Pipe ID : 2
99 | Output Schema :
100 |     n.n_nationkey : Int
101 |     n.n_name : String
102 |     n.n_regionkey : Int
103 |     n.n_comment : String
104 |     r.r_regionkey : Int
105 |     r.r_name : String
106 |     r.r_comment : String
107 | Join CNF :
108 | ( Att 2 from left record = Att 0 from right record (Int)) AND
109 | ( Att 1 from left record = Att 0 from literal record (String))
110 | *****
111 | *****
112 | Sum Operation
113 | Input Pipe ID : 2
114 | Output Pipe ID : 4
115 | Function :
116 | *****
```

Test Case 4:

Input - "SELECT SUM (n.n_regionkey) FROM nation AS n, region AS r WHERE (n.n_regionkey = r.r_regionkey) AND (n.n_name = 'UNITED STATES') GROUP BY n.n_regionkey"

```
a4-2test > F output42.txt
117 *****
118 TC4
119 *****
120 Select File Operation
121 Output Pipe ID 1
122 Output Schema:
123     n.n_nationkey : Int
124     n.n_name : String
125     n.n_regionkey : Int
126     n.n_comment : String
127 Select CNF:
128 ( Att 1 from left record = Att 0 from literal record (String))
129 *****
130 *****
131 Select File Operation
132 Output Pipe ID 3
133 Output Schema:
134     r.r_regionkey : Int
135     r.r_name : String
136     r.r_comment : String
137 Select CNF:
138
139 *****
140 *****
141 Join Operation
142 Input Pipe 1 ID : 1
143 Input Pipe 2 ID : 3
144 Output Pipe ID : 2
145 Output Schema :
146     n.n_nationkey : Int
147     n.n_name : String
148     n.n_regionkey : Int
149     n.n_comment : String
150     r.r_regionkey : Int
151     r.r_name : String
152     r.r_comment : String
153 Join CNF :
154 ( Att 2 from left record = Att 0 from right record (Int)) AND
155 ( Att 1 from left record = Att 0 from literal record (String))
156 *****
157 *****
158 Group By Operation
159 Input Pipe ID : 2
160 Output Pipe ID : 4
161 Output Schema :
162     sum : Int
163     n.n_nationkey : Int
164     n.n_name : String
165     n.n_regionkey : Int
166     n.n_comment : String
167     r.r_regionkey : Int
168     r.r_name : String
169     r.r_comment : String
170 Function :
171 OrderMaker :
172 NumAtts = 1
173 0: 3 Int
174 *****
```

Test Case 5:

Input - "SELECT SUM DISTINCT (n.n_nationkey + r.r_regionkey) FROM nation AS n, region AS r, customer AS c WHERE (n.n_regionkey = r.r_regionkey) AND (n.n_nationkey = c.c_nationkey) AND (n.n_nationkey > 10) GROUP BY r.r_regionkey"

```
a4-2test > output42.txt
175 *****
176 TC5
177 *****
178 Select File Operation
179 Output Pipe ID 1
180 Output Schema:
181     n.n_nationkey : Int
182     n.n_name : String
183     n.n_regionkey : Int
184     n.n_comment : String
185 Select CNF:
186 ( Att 0 from left record > Att 0 from literal record (Int))
187 *****
188 *****
189 Select File Operation
190 Output Pipe ID 3
191 Output Schema:
192     r.r_regionkey : Int
193     r.r_name : String
194     r.r_comment : String
195 Select CNF:
196 *****
197 *****
198 *****
199 Join Operation
200 Input Pipe 1 ID : 1
201 Input Pipe 2 ID : 3
202 Output Pipe ID : 2
203 Output Schema :
204     n.n_nationkey : Int
205     n.n_name : String
206     n.n_regionkey : Int
207     n.n_comment : String
208     r.r_regionkey : Int
209     r.r_name : String
210     r.r_comment : String
211 Join CNF :
212 ( Att 2 from left record = Att 0 from right record (Int)) AND
213 ( Att 0 from left record > Att 0 from literal record (Int))
214 *****
215 *****
216 Select File Operation
217 Output Pipe ID 4
218 Output Schema:
219     c.c_custkey : Int
220     c.c_name : String
221     c.c_address : String
222     c.c_nationkey : Int
223     c.c_phone : String
224     c.c_acctbal : Double
225     c.c_mktsegment : String
226     c.c_comment : String
227 Select CNF:
228 *****
229 *****
```



```

a4-2test > output42.txt
227 Select CNr:
228
229 *****
230 *****
231 Join Operation
232 Input Pipe 1 ID : 2
233 Input Pipe 2 ID : 4
234 Output Pipe ID : 5
235 Output Schema :
236     n.n_nationkey : Int
237     n.n_name : String
238     n.n_regionkey : Int
239     n.n_comment : String
240     r.r_regionkey : Int
241     r.r_name : String
242     r.r_comment : String
243     c.c_custkey : Int
244     c.c_name : String
245     c.c_address : String
246     c.c_nationkey : Int
247     c.c_phone : String
248     c.c_acctbal : Double
249     c.c_mktsegment : String
250     c.c_comment : String
251 Join CNF :
252 ( Att 2 from left record = Att 4 from left record (Int)) AND
253 ( Att 0 from left record = Att 3 from right record (Int)) AND
254 ( Att 0 from left record > Att 0 from literal record (Int))
255 *****
256 *****
257 Duplication Elimination Operation
258 Input Pipe ID : 5
259 Output Pipe ID : 6
260 *****
261 *****
262 Group By Operation
263 Input Pipe ID : 6
264 Output Pipe ID : 7
265 Output Schema :
266     sum : Int
267     n.n_nationkey : Int
268     n.n_name : String
269     n.n_regionkey : Int
270     n.n_comment : String
271     r.r_regionkey : Int
272     r.r_name : String
273     r.r_comment : String
274     c.c_custkey : Int
275     c.c_name : String
276     c.c_address : String
277     c.c_nationkey : Int
278     c.c_phone : String
279     c.c_acctbal : Double
280     c.c_mktsegment : String
281     c.c_comment : String
282 Function :
283 OrderMaker :
284 NumAtts = 1
285     0: 5 Int
286 *****
287 *****
288

```

Results for gTests:

```
sanjay@sanjay-VirtualBox:~/Documents/Database-Implementation/a4-2test$ ./gtest.out
[=====] Running 5 tests from 1 test suite.
[-----] Global test environment set-up.
[-----] 5 tests from EXECUTIONPLAN
[ RUN      ] EXECUTIONPLAN.TC1
[ OK       ] EXECUTIONPLAN.TC1 (4 ms)
[ RUN      ] EXECUTIONPLAN.TC2
[ OK       ] EXECUTIONPLAN.TC2 (1 ms)
[ RUN      ] EXECUTIONPLAN.TC3
[ OK       ] EXECUTIONPLAN.TC3 (1 ms)
[ RUN      ] EXECUTIONPLAN.TC4
[ OK       ] EXECUTIONPLAN.TC4 (1 ms)
[ RUN      ] EXECUTIONPLAN.TC5
[ OK       ] EXECUTIONPLAN.TC5 (2 ms)
[-----] 5 tests from EXECUTIONPLAN (19 ms total)

[-----] Global test environment tear-down
[=====] 5 tests from 1 test suite ran. (21 ms total)
[ PASSED  ] 5 tests.
sanjay@sanjay-VirtualBox:~/Documents/Database-Implementation/a4-2test$ |
```