Project Report 3 Implementing Relational Operators

Arpan Banerjee UFID: 9359-9083 arpanbanerjee@ufl.edu

Krutantak Patil UFID: 5615-6343 Krutantakb.patil@ufl.edu

1) Steps to compile and run - code, tests and gtests

(Assuming .tbl files would be provided to us in same directory)

- i. make a2test.out Command to compile a2test.cc which creates bin files.
- ii. ./a2test.out Command to run a2test.cc.
- iii. make test.out Command to build the tests.
- iv. ./test.out Command to run the tests.
- v. **make gtest.out** Command to build the gtests.
- vi. ./gtest.out Command to run the gtests.

2) Implementations of Relational Operators

SelectFile

i. Data members

- a. pthread_t thread Thread started by Run function.
- b. Record *buffer
- c. DBFile *inFile Pointer to input DBFile
- d. Pipe *outPipe Pointer to output pipe
- e. CNF *selOp Pointer to CNF used to select record
- f. Record *literal

ii. Run (DBFile &inFile, Pipe &outPipe, CNF &selOp, Record &literal)

- a. Used to set the data member pointers and start a thread using pthread_create for processing.
- b. This function does the same task for all the operations.

iii. WaitUntilDone ()

- a. Used to wait for and join the background thread created in Run.
- b. This is a blocking function and performs the same task for all operations.

iv. **Use_n_Pages** (int runlen)

- a. In Operations that use a BigQ instance, this function is used to initialize the run length. For other operations, it does nothing.
- b. For SelectFile, it does nothing.

v. select_file_function ()

- a. This function runs on the thread created by Run.
- b. It uses GetNext with the CNF argument to get records from the DBFile and inserts them into the output pipe.

SelectPipe

i. Data members

- a. pthread_t thread
- b. Record *buffer
- c. Pipe *inPipe
- d. Pipe *outPipe
- e. CNF *selOp
- f. Record *literal

ii. select_pipe_function ()

- a. This function runs on the thread created by Run.
- b. It removes records one by one from the input pipe, compares them using the CNF and inserts it into the output pipe if it's a match.

Project

i. Data members

- a. pthread_t thread
- b. Pipe *inPipe
- c. Pipe *outPipe
- d. int *keepMe Integer array of indices of the record that we want to keep.
- e. int numAttsInput number of attributes in input records
- f. int numAttsOutput number of output attributes to project to.
- g. Record *buffer

ii. project_function ()

- a. This function runs on the thread created by Run.
- b. It removes records one by one from the input pipe, uses Project() and then inserts it into the output pipe.

WriteOut

i. Data members

- a. pthread_t thread
- b. Record *buffer
- c. Pipe *inPipe
- d. FILE *outFile File to write to.
- e. Schema *mySchema Schema of the records to write out.

ii. writeout_function ()

a. This function runs on the thread created by Run.

b. It removes records one by one from the input pipe, goes through all its attributes and keeps writing it to the file.

DuplicateRemoval

i. Data members

- a. pthread t thread
- b. Pipe *inPipe
- c. Pipe *outPipe
- d. Schema *mySchema
- e. int runlen

ii. duplicate_function ()

- a. This function runs on the thread created by Run.
- b. First, the records are sorted (based on all its attributes) using a BigQ.
- c. Then we iterate through the sorted records and insert into the output pipe when the new record is not same as the previous record.

Sum

i. Data members

- a. pthread t thread
- b. Pipe *inPipe
- c. Pipe *outPipe
- d. Function *func Function to apply on the record to calculate sum.
- add_result (Type rtype, int &isum, int ires, double &dsum, double dres)
 - a. This utility function adds the result into the int sum or the double sum depending on the return type.

iii. **sum_function** ()

- a. This function runs on the thread created by Run.
- b. Records are removed one by one from the input pipe and the function is applied. The results are accumulated using add_result.
- c. Then we create a new schema with 1 attribute (sum) to put the result record into the output pipe.

GroupBy

i. Data members

- a. pthread_t thread
- b. int runlen

- c. Pipe *inPipe
- d. Pipe *outPipe
- e. OrderMaker *groupAtts
- f. Function *func

ii. PerformGroupBy ()

- a. This function runs on the thread created by Run.
- b. First, the records are sorted using a BigQ and the attributes that we need to group by.
- c. Then, until the value of the record changes (wrt group attributes) we keep adding the value of the function and insert it into the output pipe for each separate group.

Join

i. Data members

- a. CNF *selOp
- b. int runLength
- c. Pipe *leftInPipe, *rightInPipe, *outPipe
- d. pthread_t joinThread
- e. Record *literal

ii. PerformJoin ()

- a. This function runs on the thread created by Run.
- b. Two instances of BigQ are used, one for the left pipe and one for the right pipe.
- c. These are then merged using the algorithm described in the problem statement or a nested block loop.

3) ./test.out results for 1GB data

i. Q1

select * from partsupp where ps_supplycost <1.03

```
**IMPORTANT: MAKE SURE THE INFORMATION BELOWITS COMMETC) **String: [orange floral clive ivory lace], double: [931.03] technique to the control of the contro
```

ii. Q2

select p_partkey(0), p_name(1), p_retailprice(7) from part where (p_retailprice > 931.01) AND (p_retailprice < 931.3);

iii. Q3

select sum (s acctbal + (s acctbal * 1.05)) from supplier;

```
Shadow-reaper-u-vm@shadow-reaper-u-vm:~/Shadow_geaper-U-Vm/Projects/database-system-implementation/a3test$ ./test.out 3

### IMPORTANTORMAKE SURE THE INFORMATION BELOW IS CORRECT **

catalog clocation: catalog

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### CRECOPIC Albert

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT

### TIMPORTANT
```

iv. Q4

select sum (ps_supplycost) from supplier, partsupp where s suppkey;

v. Q5

select distinct ps suppkey from partsupp where ps supplycost < 100.11;

vi. Q6

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

```
** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog clocation: catalog 35
tpch files dir: ... 37
heap files dir: ... 39
Catalog clocation: catalog 35
tpch files dir: ... 37
heap files dir: ... 37
The Relop catalog 10 catalog 1
```

In q6, we have modified the test to initialize the ordermaker as it wasn't working initially.

3) GTests and results

- **i. SELECT_GROUP_BY**: We join the suppliers and partsupp tables using "s_suppkey = ps_suppkey". Then we group it by ps_supplycost. Finally we assert our output has 25 records.
- **ii. SELECT_PIPE_1:** We select from the partsupp table "ps_suppcost < 1.03" and verify that we get 21 records.
- **SELECT_PIPE_2:** We select from the partsupp table "ps_suppcost < 1.04" and verify that we get 31 records.
- **iv. PROJECT:** We select "(p_retailprice > 931.00) AND (p_retailprice < 931.4)" from parts table, then project it to keep attributes 0, 1 and 7.
- **V. SUM_TEST:** From nation table, select using "n_regionkey > 3" and sum n_regionkey.

```
double: [1.27399e-313]
double: [1.48618e-313]
double: [1.69842e-313]
  ouble:DATADA9106EM3134MENTATION
 double: =[2:12277e-313]
double: [2,54714e-313]
double: [2,54714e-313]
double: [2,75931e-313]
double: [2,97154e-313]
double: [3,18373e-313]
 double: ≣[3,396e-313]
 double: <u>=</u>[3.60822e-313]
 double: [3.8204e-313]
double: [4.03258e-313]
double: [4.24481e-313]
double: [4.45698e-313]
double: [4.66919e-313]
 double: =[4:88136e-313]
 double: [5.09357e-313]
                OK J RELATIONAL OPERATION. SELECT GROUP BY (13709 ms)
                E part bin metal 2 records 68 query2 returned 12 records
  * IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
  catalog location:
                                                    catalog
  tpch files dir:
  heap files dir:
 FILE TYPE 0
 PAGE COUNT 1065
os partkey: [6333], ps_suppkey: [6334], ps_availqty: [3711], ps_supplycost: [1.01],
ps_partkey: [9097], ps_suppkey: [4098], ps_availqty: [3012], ps_supplycost: [1.01],
al packages above the unusual, bold packages cajole blithely even Tiresias. theodoli
ps_partkeyeor[20468], ps_suppkey: [469]GE psulvallqty: [6884], ps_supplycost: [1], ps
ounts cajole deposits above the ij double: [9.24623e+07]
ps_partkey: [27115], ps_suppkey: [9618], ps_availqty: [7966], ps_supplycost: [1.02], ress requests. deposits unwind carefully pending theodolites. pinto beans about the ps_partkey: [34494], ps_suppkey: [9501], ps_availqty: [7438], ps_supplycost: [1.02], ps_partkey: [43172], ps_suppkey: [685], ps_availqty: [6600], ps_supplycost: [1.01], lar, even accounts haggle carefully slyly bold requests. even pinto beans ] ps_partkeyn: [43764], ps_suppkey: [1277], ps_availqty: [2344], ps_supplycost: [1.02], pipto beans_use blithely careful]
```

```
TLE TYPE Page Func.
      unts cajote, deposits above the 1]
s partKey: [27115], ps_suppkey: [9618], ps_availqty: [7966], ps_supplycost: [1.02], ps_comment: [e regular, iro
ess requests: deposits unwind carefully pending theodolites, pinto beans about the even, regular theodolite]
s partKey: [29444], ps_suppkey: [9445], ps_availqty: [3118], ps_supplycost: [1.03], ps_comment: [l requests thr
ake slyly, according]
    The style according in the supple of the sup
     be partkey: [193659], ps suppkey: [6179], ps availaty: [6606], ps supplycost: [1.01], ps comment: [ca
bunts steep special packages. ironic ideas according to the furiously regular dolphins use quickly pl
     * IMPORTANT FMAKE SURE THE INFORMATION BELOW IS CORRECT **
      catalog location: cata
tpch filessdiff MARLEMENTATION
       heap files dim
    FILE TYPE O On bin
PAGE COUNT 1965 meta 47 int: [1930], string: [orange floral olive ivory lace], double: [931.03] int: [31] string: [slate seashell steel medium moceasining property along the string: [orange floral olive 1920] int: [1030] string: [orange floral olive 1920] int: [1930] string: [orange floral olive 1920] int: [2029], string: [middight brown olim 0100] to 1920] int: [2029], string: [puff slate tomato moceasin 1920] int: [3028], string: [puff slate tomato moceasin 1920] int: [3028], string: [white ivory moceasin 1920] int: [3028], string: [white ivory moceasin 1920] int: [3027], string: [white ivory moceasin 1920] int: [3028], string: [blanched blush brink 1230] int: [5026] 
      PAGE COUNT 1262 n.meta
                                                                                                                     [navajo tan almond yettow blue], "double: [931.01]
[brown royal steel forest chiffon], double: [931.01]
[blue chiffon green powder burnished], double: [931.01]
[ghost aquamarine spring coral lace], double: [931.01]
[forest lavender lime antique medium], double: [931.01]
[rosy yellow spring wheat orange], double: [931.01]
[drab linen blue almond cream], double: [931.01]
                            [13018], string:
   int: [14017], string:
   int: [15016], string:
int: [16015], string:
                            [17014] string:
[18013], string:
    int:
```

```
[ RUN | RELATIONAL OPERATION. PROJECT (130 ms) | RELATIONAL OPERATION. SUMSTEST returned 12 records | RELATIONAL OPERATION. SUMSTEST returned 12 records | RELATIONAL OPERATION. SUMSTEST returned 12 records | Relation | R
```