COP 6726: DATABASE SYSTEM IMPLEMENTATION Assignment 2.1: Sorted File Implementation

Group Info:

Name UFID

Manishkumar Chopra
 Neel Manish Rami
 7712-3151

I. INTRODUCTION

In this assignment, we have implemented the BigQ class. The BigQ class encapsulates the process of taking a stream of inserts, breaking the stream of inserts into runs, and then using an in-memory priority queue to organize the head of each run and give the records to the caller in sorted order. We have also written unit test cases using google test unit testing library for every function which we have implemented in DBFile class.

II. ASSUMPTIONS

- A. gcc (or clang) must be installed together with "bison" and "flex" on the machine on which the project will be tested.
- B. Google Test Unit Testing Library should also be installed and compiled on the machine on which the project will be tested.
- C. Google Test Unit Testing Library is not installed on CISE Thunder Server so the gTest command won't run on thunder.

III. FILE STRUCTURE

- A. bin/: Contains object(.o) files.
- B. src/: Contains .cc and .h files.
- C. table_files/: Contains .tbl files which were generated using tpch data generator.
- D. gtest/: Google Test Unit testing library.
- E. catalog: Catalog file for the schema of .tbl files
- F. MakeFile: Make File.

IV. CODE EXPLANATION

Data Structures used in BigQ Class:-

- 1. **struct TPMMS_DS:** This structure is used during the algorithm.
- struct RecordSorter: This struct is used by the vector of Records to sort records.
- 3. struct RecordComparator: This struct is used by a priority queue.
- **4. struct Record1:** It's a wrapper around Record Class which has an instance of Record Class and an integer which stores the run number to which the record belongs.

Public Functions used in BigQ Class:-

1. void* TPMMS_Algo(void* args)

This method implements the TPMMS Algorithm/

V. COMMANDS TO RUN THE PROGRAM

To compile the given test driver, type **make test.out**

To run the test driver, type

./test.out

and then follow the on-screen instructions. For scan and filter operation, write any CNF which has been given in the result section below. After writing the CNF, press ENTER and then press Ctrl+D.

The thunder server at cise.ufl.edu doesn't have Google Test installed so you might have to download it.

To compile Google Test, type make gTest2

To run the unit tests, type cd bin ./gTestProi2

The unit tests should run and you should be able to see the test results.

VI. RESULTS

We ran our program on 10M dataset.

A. 10M Data

We verified our results with the results which have been provided by the professor in the assignment folder.

Query 1 (Sort): (r_name)

```
## thunder:sash /test.out

**IMMONTANI, NAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location: catalog
catalog in catalog
specified dir: ./table_files/
heap files dir: ./table_files/
heap files dir: ./table_files/
heap files dir: ./table_files/
select dbfile to use:

1. select dbfile to use:

2. region
2. region
3. sort + Write
1. select dbfile to use:

1. part
2. region
3. part supp
6. orders
6. orders
7. linelium
2. specify sort ordering (when done press ctrl-D):
(c_name)
producer: opened BBfile region.bin
producer: inserted 5 sees into the pipe
consumer; 5 rees out of 5 rees in sorted order
thunder:408

Type here to search

1. Type here to search
```

Query 2 (Sort): (n_name)

```
# IMPORTANT: MARE SURE THE INFORMATION BELOW IS CORRECT **

CITATION ACCURATE THE INFORMATION BELOW IS CALLED THE INFORMATION BELOW IS CALLED
```

Query 3 (Sort):

(o_totalprice)

```
# Ununder read the du - PLITY

Thunder: 47% - //east.out

Thunder: 47% - //east.out

The MONOTRANN, MANK SUBE THE INFORMATION BELOW IS CORRECT **

catalog location:

catalog catalog

the files dis:

/table_files/
heap files dis:

1. sort

2. sort + display

3. sort + write

1

select defile to use:

1. nation

2. region

3. customer

4. purtupp

6. orders

7. lineitem

6

specify sort ordering (when done press ctrl-D):

(_otalprice)

producer: opened DBFile coders.bin

producer: opened DBFile coders.bin

consumer: removed 15000 recs into the pipe

consumer: removed 15000 recs from the pipe

consumer: removed 15000 recs from the pipe

consumer: removed 15000 recs from the pipe

consumer: removed 15000 recs out of 15000 recs in sorted order

thunder: 48% 

O Type here to search

D Type here to search

C Type here to s
```

Query 4: (Sort)

(ps_suppkey) AND (ps_partkey)

```
## Unuder.issafledu-PuTY

thunder:45* /test.out

** INDORTANT, NAME SURE THE INFORMATION BELOW IS CORRECT **
catalog location: catalog
tech files dir: /table_files/
heap files dir: /table_files/
heap files dir:

1. mort
2. sort + display
3. sort + write
1

select defile to use:
1. nation
2. region
3. oustower
4. purt
4. purt
5. orders
7. lineitem
5

specify runlength:
32

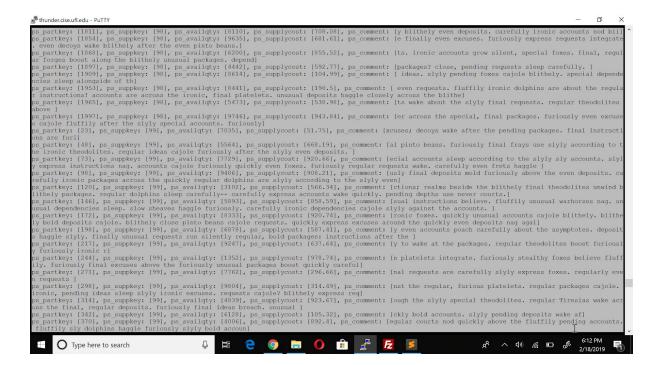
specify sort ordering (when done press ctrl-D):
(ps_suppkey) AND (ps_partkey)
producer: opened DBFile partsupp.hin
producer: inserted 8000 recs from the pipe
consumer: removed 8000 recs from the pipe
consumer: season and the pipe
consumer: removed 8000 recs from the pipe
consumer: season and the pipe
consumer: season and the pipe
consumer: removed 8000 recs from the pipe
consumer: season and the pipe
consumer: season and the pipe
consumer: removed 8000 recs from the pipe
consumer: season and the pipe
consumers and the p
```

Query 5 (Sort + Display):

(n_name)

Query 6 (Sort + Display):

(ps_suppkey) AND (ps_partkey)



Query 7 (Sort + Display):

(l_shipdate) AND (l_extendedprice) AND (l_quantity)

```
REG ARR], ]_comment: [n, even Warthogs akes]

REG ARR], ]_comment: [n, express packages wate. furiously re]

Lorderkey: [1953], ]_partxey: [315], l. jarkpeky: [16], l. linemuber: [1], l. quantity: [31], l. extendedprice: [37674.6], l. discount: [0.03], l. tax: [0.02], l. return[lag; [18], l. comment: [deposits happigle about [1], linemuber: [1], l. quantity: [31], l. extendedprice: [1998-12-11], l. phipinstruct: [7MNE BACK EXTURN], l. return[lag; [18], l. linestatus: [0], l. phipate: [1998-11-22], l. committate: [1998-09-20], l. receiptdate: [1998-12-16], l. phipinstruct: [7MNE BACK EXTURN], l. return[lag; [18], l. linestatus: [0], l. phipate: [1998-11-22], l. committate: [1998-09-20], l. receiptdate: [1998-12-16], l. phipinstruct: [7MNE BACK EXTURN], l. return[lag; [18], l. linestatus: [0], l. phipate: [1998-11-22], l. committate: [1998-09-20], l. receiptdate: [1998-12-16], l. phipinstruct: [7MNE BACK EXTURN], l. discount: [1984], l. partkey: [366], l. suppkey: [55], l. linenumber: [1], l. quantity: [33], l. extendedprice: [4798-9], l. discount: [0.11], l. tax: [0.01], l. phipate: [1998-11-23], l. committate: [1998-11-24], l. phipate: [1998-11-24], l. p
```

Query 8 (Sort + Display):

(r_comment)

```
## Comment | Com
```

Query 9 (Sort + Write):

(o_totalprice) AND (o_shippriority)

Query 10 (Sort + Write):

(p_size) AND (p_type)

```
C_Cotalprice AND (c_shippriority)

producer; opened DBFile orders.bin
producer; inserted 15000 recs into the pipe
consumer: recoved 15000 recs from the pipe
consumer: recover inserted 15000 recs into the pipe
consumer: recover inserted 15000 recs of 15000 recs in sorted order
thunder: f4% ./test.out

** MINORTAMY, MMAS SUBE THE INFORMATION BELOW IS CORRECT **
catalog location: catalog
the files dir: ./table_files/
heap files dir: ./table_files/
heap files dir: ./table_files/

** select test option:
1. sort
2. sort ** display
3. sort ** write
3

** select thiftie to use:
1. nation
2. region
3. customer
4. partump
6. orders
7. limitem
4

**specify runlength:
4

**specify runlength:
4

**specify runlength: catalog translation
producer: inserted 2000 recs from the pipe
consumer: recover 2000 recs out of 2000 recs in sorted order
thunder: f5%

**Different And The Add T
```

Google Test Results: