

PROJECT REPORT
IMPLEMENTING A SORTED FILE
MILESTONE 1:
THE BIGQ CLASS

Name: Gloria Gilbert Nazareth
UFID: 8221-8035
GatorLink: gnazareth@ufl.edu

Steps to compile and execute:

1. make ./test.out : this command builds the program and creates the required object files
2. ./test.out: this will execute the program, and ask for user input in order to perform required functionality.
3. ./runTestCases.sh: this will run the test cases mentioned in the file runTestCases.sh and will write the output to output1.txt
4. Make ./gtest.out : this command will build the program and create the required object files
5. ./runGtest.sh : this will execute the gtest. We are executing it through a script file, as the yy_parse() function used to test sorting and priority queue requires user input.
[PS: Bin files need to be present before compiling the code, else it will result in error]

Description of the Implemented Function:

The BigQ class has the following functions:

1. BigQ (Pipe &in, Pipe &out, OrderMaker &sortorder, int runlen)
This is the constructor of the class. It declares a worker pthread and starts it. It also passes the workerRoutine to the pthread to perform the necessary Two Pass Multiple Merge (TPMM) algorithm. The constructor terminates if the runlen is zero.
2. void* WorkerRoutine(void* arg)
This function is called when the pthread is declared. This procedure is responsible for the TPMM algorithm. The function calls two other functions divideRuns and mergeRuns needed for the algorithm.
3. void divideRuns(Pipe *in, int runlen, map<int,Page*>& overflow, OrderMaker* sortorder, File& f, int& pageCount)
This function reads the records from the in pipe and writes it to pages. Once a page is full, it appends it to the page vector by calling the addRecords function. It also counts the number of pages written. Once the number of pages written is equal to runlength (runlen) it calls the sortRunlens method and then writes the sorted pages in the runlen to a temporary file.
4. void addRecords(vector<Record*>& recVector, vector<Page*>& pageVector)
This function adds the records from the pages in the pageVector to the recordVector.
5. int sortRunlens(vector<Record*>& recVector, OrderMaker* sortorder)
This function sorts the records present in the record vector using the sortorder comparator.
6. int writeSortedRun(map<int,Page*>& overflow, File& f, vector<Record*> recVector, int runlen, int& pageCount)
This function writes the sorted records in the record vector onto pages and appends it to a file. It also checks for overflows and maps those pages in a overflow map, which is used in merging.

7. void mergeRuns(int pageCount, Pipe* out, int runlen, File& f, OrderMaker* sortorder, map<int,Page*> overflow)

This function merges the sorted runlens to create a sorted file. It does this using a priority queue. The algorithm used is similar to K-way merging of lists. It calculates the number of runs by dividing the total pages by run length. A Map (recRunIndexMap) is used to map the records to a run number. An array (pageArray) is used to map the current page number to a run number. After all the records are sorted, i.e the queue is empty, the out pipe is shutdown.

Output:

10 Mb data:

Console:

```
gloria@gloria-inspiron-5593: /media/loria/Work/UniversityofFlorida/Study/SemII/OBI/Projects/Project2/PartI/a2test$ ./runTestCases.sh
consumer: 25 recs out of 25 recs in sorted order
consumer: recs removed written out as heap DBFile at bin/customer.bin.bigq
consumer: 1500 recs out of 1500 recs in sorted order
consumer: 15000 recs out of 15000 recs in sorted order
consumer: 25 recs out of 25 recs in sorted order
```

Output1.txt:

```
output1.txt (Work /media/loria/Work/UniversityofFlorida/...udy/SemII/OBI/Projects/Project2/PartI/a2test) - gedit
n_nationkey: [6], n_name: [FRANCE], n_regionkey: [3], n_comment: [refully final requests. regular, iron]
n_nationkey: [19], n_name: [ROMANIA], n_regionkey: [3], n_comment: [ular asymptotes are about the furious multipliers. express dependencies nag above the ironically ironic account]
n_nationkey: [22], n_name: [RUSSIA], n_regionkey: [3], n_comment: [ requests against the platelets use never according to the quickly regular pint]
n_nationkey: [23], n_name: [UNITED KINGDOM], n_regionkey: [3], n_comment: [eans boost carefully special requests. accounts are. careful]
n_nationkey: [13], n_name: [JORDAN], n_regionkey: [4], n_comment: [tc deposits are blithely about the carefully regular pa]
n_nationkey: [4], n_name: [EGYPT], n_regionkey: [4], n_comment: [y above the carefully unusual theodolites. final dugouts are quickly across the furiously regular d]
n_nationkey: [11], n_name: [IRAQ], n_regionkey: [4], n_comment: [nic deposits boost atop the quickly final requests! quickly regula]
n_nationkey: [20], n_name: [SAUDI ARABIA], n_regionkey: [4], n_comment: [ts. silent requests haggle. closely express packages sleep across the blithely]
n_nationkey: [10], n_name: [IRAN], n_regionkey: [4], n_comment: [efully alongside of the slyly final dependencies. ]
consumer: removed 25 recs from the pipe
*****
File bin/customer.bin.bigq created successfully
producer: inserted 1500 recs into the pipe
consumer: removed 1500 recs from the pipe
*****
o_orderkey: [39620], o_custkey: [1279], o_orderstatus: [F], o_totalprice: [406930], o_orderdate: [1994-10-05], o_orderpriority: [2-HIGH], o_clerk: [Clerk#000000457], o_shippriority: [0], o_comment: [nic
packages sleep quickly regular excuses. busy packages nag finally! package]
o_orderkey: [17571], o_custkey: [358], o_orderstatus: [F], o_totalprice: [408346], o_orderdate: [1992-03-16], o_orderpriority: [4-NOT SPECIFIED], o_clerk: [Clerk#0000000503], o_shippriority: [0],
o_comment: [ckages. blithely silent requests use quickly final packages. somas mainta]
o_orderkey: [39456], o_custkey: [1016], o_orderstatus: [0], o_totalprice: [409771], o_orderdate: [1998-02-16], o_orderpriority: [2-HIGH], o_clerk: [Clerk#000000972], o_shippriority: [0], o_comment: [ts
sleep final requests. carefully b]
o_orderkey: [57376], o_custkey: [676], o_orderstatus: [0], o_totalprice: [411255], o_orderdate: [1995-06-20], o_orderpriority: [1-URGENT], o_clerk: [Clerk#000000618], o_shippriority: [0], o_comment: [ular
tiresias after the furiously regular requests sha]
o_orderkey: [6882], o_custkey: [178], o_orderstatus: [0], o_totalprice: [422360], o_orderdate: [1997-04-09], o_orderpriority: [1-URGENT], o_clerk: [Clerk#000000974], o_shippriority: [0], o_comment: [y
dolphins. furiously ironic asymptotes s]
o_orderkey: [59106], o_custkey: [953], o_orderstatus: [0], o_totalprice: [430620], o_orderdate: [1996-10-24], o_orderpriority: [4-NOT SPECIFIED], o_clerk: [Clerk#0000000344], o_shippriority: [0],
o_comment: [y above the slyly final deposits. furiously regular pinto beans use s]
o_orderkey: [44707], o_custkey: [1013], o_orderstatus: [0], o_totalprice: [431772], o_orderdate: [1997-08-14], o_orderpriority: [1-URGENT], o_clerk: [Clerk#0000000313], o_shippriority: [0], o_comment: [ly
fluffly special ideas.]
o_orderkey: [29150], o_custkey: [667], o_orderstatus: [0], o_totalprice: [439687], o_orderdate: [1995-10-21], o_orderpriority: [2-HIGH], o_clerk: [Clerk#000000525], o_shippriority: [0], o_comment: [lessly
quiet deposits across the furiously iron]
o_orderkey: [52965], o_custkey: [676], o_orderstatus: [0], o_totalprice: [446001], o_orderdate: [1996-09-22], o_orderpriority: [3-MEDIUM], o_clerk: [Clerk#000000103], o_shippriority: [0], o_comment:
[thless pinto beans use about the blithe]
consumer: removed 1500 recs from the pipe
*****
n_nationkey: [7], n_name: [GERMANY], n_regionkey: [3], n_comment: [l platelets. regular accounts x-ray: unusual, regular acco]
n_nationkey: [19], n_name: [ROMANIA], n_regionkey: [3], n_comment: [ular asymptotes are about the furious multipliers. express dependencies nag above the ironically ironic account]
n_nationkey: [22], n_name: [RUSSIA], n_regionkey: [3], n_comment: [ requests against the platelets use never according to the quickly regular pint]
n_nationkey: [23], n_name: [UNITED KINGDOM], n_regionkey: [3], n_comment: [eans boost carefully special requests. accounts are. careful]
n_nationkey: [4], n_name: [EGYPT], n_regionkey: [4], n_comment: [y above the carefully unusual theodolites. final dugouts are quickly across the furiously regular d]
n_nationkey: [10], n_name: [IRAN], n_regionkey: [4], n_comment: [efully alongside of the slyly final dependencies. ]
n_nationkey: [11], n_name: [IRAQ], n_regionkey: [4], n_comment: [nic deposits boost atop the quickly final requests! quickly regula]
n_nationkey: [13], n_name: [JORDAN], n_regionkey: [4], n_comment: [tc deposits are blithely about the carefully regular pa]
n_nationkey: [20], n_name: [SAUDI ARABIA], n_regionkey: [4], n_comment: [ts. silent requests haggle. closely express packages sleep across the blithely]
consumer: removed 25 recs from the pipe
```

1 GB data:

Console:

```
gloria@gloria-Inspiron-5593:/media/gloria/Work/UniversityOfflorida/Study/SemII/DBI/Projects/Project2/Part1/a2test$ ./runTestCases.sh
consumer: 25 recs out of 25 recs in sorted order
producer: 100000
consumer: recs removed written out as heap DBFile at bin/customer.bin.bigq
consumer: 150000 recs out of 150000 recs in sorted order
producer: 100000
producer: 200000
producer: 300000
producer: 400000
producer: 500000
producer: 600000
producer: 700000
producer: 800000
producer: 900000
producer: 1000000
producer: 1100000
producer: 1200000
producer: 1300000
producer: 1400000
producer: 1500000
consumer: 1500000 recs out of 1500000 recs in sorted order
consumer: 25 recs out of 25 recs in sorted order
gloria@gloria-Inspiron-5593:/media/gloria/Work/UniversityOfflorida/Study/SemII/DBI/Projects/Project2/Part1/a2test$
```

Output1.txt:

```
output1.txt (Work /media/gloria/Work/UniversityOfflorida/...udy/SemII/DBI/Projects/Project2/Part1/a2test) - gedit
n_nationkey: [6], n_name: [FRANCE], n_regionkey: [3], n_comment: [refully final requests. regular, lronl]
n_nationkey: [19], n_name: [ROMANIA], n_regionkey: [3], n_comment: [ular asymptotes are about the furious multipliers, express dependencies nag above the ironically ironic account]
n_nationkey: [22], n_name: [RUSSIA], n_regionkey: [3], n_comment: [ requests against the platelets use never according to the quickly regular pint]
n_nationkey: [23], n_name: [UNITED KINGDOM], n_regionkey: [3], n_comment: [eans boost carefully special requests. accounts are. careful]
n_nationkey: [13], n_name: [JORDAN], n_regionkey: [4], n_comment: [ic deposits are blithely about the carefully regular pa]
n_nationkey: [4], n_name: [EGYPT], n_regionkey: [4], n_comment: [y above the carefully unusual theodolites. final dupouts are quickly across the furiously regular d]
n_nationkey: [11], n_name: [IRAQ], n_regionkey: [4], n_comment: [nlc deposits boost atop the quickly final requests? quickly regula]
n_nationkey: [20], n_name: [SAUDI ARABIA], n_regionkey: [4], n_comment: [ts. silent requests haggle. closely express packages sleep across the blithely]
n_nationkey: [10], n_name: [IRAN], n_regionkey: [4], n_comment: [efully alongside of the slyly final dependencies. ]
consumer: removed 25 recs from the pipe
=====
file bin/customer.bin.bigq created successfully
producer: inserted 150000 recs into the pipe
consumer: removed 150000 recs from the pipe
=====
o_orderkey: [4515876], o_custkey: [1006085], o_orderstatus: [F], o_totalprice: [$10062], o_orderdate: [1993-11-02], o_orderpriority: [4-NOT SPECIFIED], o_clerk: [Clerk#000000105], o_shippriority: [0],
o_comment: [carefully accounts: slyly ironic pinto beans detect. brave sonas a]
o_orderkey: [2185667], o_custkey: [51796], o_orderstatus: [F], o_totalprice: [$11360], o_orderdate: [1992-10-00], o_orderpriority: [1-URGENT], o_clerk: [Clerk#000000574], o_shippriority: [0], o_comment:
[ deposits wake quickly unusual deposits. express dependencies wake. ]
o_orderkey: [2199712], o_custkey: [66790], o_orderstatus: [0], o_totalprice: [$15532], o_orderdate: [1996-09-30], o_orderpriority: [2-HIGH], o_clerk: [Clerk#000000650], o_shippriority: [0], o_comment:
[ the final, ironic deposits lnte]
o_orderkey: [3386919], o_custkey: [24049], o_orderstatus: [F], o_totalprice: [$22644], o_orderdate: [1992-11-07], o_orderpriority: [1-URGENT], o_clerk: [Clerk#000000924], o_shippriority: [0], o_comment:
[are alongside of the pending deposits. quick]
o_orderkey: [2232932], o_custkey: [13940], o_orderstatus: [0], o_totalprice: [$22721], o_orderdate: [1997-04-13], o_orderpriority: [2-HIGH], o_clerk: [Clerk#000000245], o_shippriority: [0], o_comment:
[litely stealthy accounts are slyly against the pend]
o_orderkey: [4576548], o_custkey: [108931], o_orderstatus: [0], o_totalprice: [$25591], o_orderdate: [1997-12-26], o_orderpriority: [1-URGENT], o_clerk: [Clerk#000000336], o_shippriority: [0], o_comment:
[he slyly ironic requests. regular, bold deposits cajol]
o_orderkey: [3043270], o_custkey: [144617], o_orderstatus: [0], o_totalprice: [$30604], o_orderdate: [1997-02-12], o_orderpriority: [5-LOW], o_clerk: [Clerk#000000699], o_shippriority: [0], o_comment:
[riously final deposits?]
o_orderkey: [4722021], o_custkey: [128120], o_orderstatus: [F], o_totalprice: [$44089], o_orderdate: [1994-04-07], o_orderpriority: [1-URGENT], o_clerk: [Clerk#000000230], o_shippriority: [0], o_comment:
[al, express pinto beans are after the careful]
o_orderkey: [1750460], o_custkey: [21433], o_orderstatus: [F], o_totalprice: [$55285], o_orderdate: [1992-11-30], o_orderpriority: [4-NOT SPECIFIED], o_clerk: [Clerk#000000400], o_shippriority: [0],
o_comment: [ ironic packages. even notornis integrate. fluffily regular foxes snooze furi]
consumer: removed 1500000 recs from the pipe
=====
n_nationkey: [7], n_name: [GERMANY], n_regionkey: [3], n_comment: [l platelets. regular accounts x-ray: unusual, regular acco]
n_nationkey: [19], n_name: [ROMANIA], n_regionkey: [3], n_comment: [ular asymptotes are about the furious multipliers, express dependencies nag above the ironically ironic account]
n_nationkey: [22], n_name: [RUSSIA], n_regionkey: [3], n_comment: [ requests against the platelets use never according to the quickly regular pint]
n_nationkey: [23], n_name: [UNITED KINGDOM], n_regionkey: [3], n_comment: [eans boost carefully special requests. accounts are. careful]
n_nationkey: [4], n_name: [EGYPT], n_regionkey: [4], n_comment: [y above the carefully unusual theodolites. final dupouts are quickly across the furiously regular d]
n_nationkey: [10], n_name: [IRAQ], n_regionkey: [4], n_comment: [efully alongside of the slyly final dependencies. ]
n_nationkey: [11], n_name: [IRAQ], n_regionkey: [4], n_comment: [nlc deposits boost atop the quickly final requests? quickly regula]
n_nationkey: [13], n_name: [JORDAN], n_regionkey: [4], n_comment: [ic deposits are blithely about the carefully regular pa]
n_nationkey: [20], n_name: [SAUDI ARABIA], n_regionkey: [4], n_comment: [ts. silent requests haggle. closely express packages sleep across the blithely]
Consumer: removed 25 recs from the pipe
```

Gtest Results:

1. TEST(writeSortedRuns,test1)

This function will read records from a file, write it to a record vector. It then tests the writeSortedRun function in BigQ class. Here we check it for a file with records less than a single page.

2. TEST(writeSortedRunsMultiple,test2)

This function will read records from a file, write it to a record vector. It then tests the writeSortedRun function in BigQ class. Here we check it for a file with records more than a single page.

3. TEST(sortRun, test3)

This function tests the sorting of a records vector. Here, we read the records from an unsorted bin file and a sorted bin file and push it to two different record vectors. We then sort the vector containing the records of the unsorted bin file using the SortRunlens function of BigQ. After this we compare the records one by one. The assert equates to 0 if unsorted record vector is equal to the sorted record vector. Else 1.

4. TEST(checkPriorityQueue, test4)

This function checks if the priority queue has the records in a sorted manner. This function tests the sorting of a records vector. Here, we read the records from an unsorted bin file and a sorted bin file and push it to two different record vectors. We then create a priority queue with a particular Merge_sort comparator. It then pops the records one by one and we cross check it with the sorted record vector. If it exactly matches it returns 0, else 1.

```
gloria@gloria-Inspiron-5593:/media/gloria/Work/UniversityOfFlorida/Study/SenII/DBI/Projects/Project2/Part1/a2test$ ./runGtest.sh

** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
catalog location:    catalog
tpch files dir:      data/tpch-dbggen/
heap files dir:      bin/

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

[*****] Running 4 tests from 4 test cases.
[-----] Global test environment set-up.
[-----] 1 test from writeSortedRuns
[ RUN ]   writeSortedRuns.test1
[ OK ]    writeSortedRuns.test1 (1 ms)
[-----] 1 test from writeSortedRuns (1 ms total)

[-----] 1 test from writeSortedRunsMultiple
[ RUN ]   writeSortedRunsMultiple.test2
[ OK ]    writeSortedRunsMultiple.test2 (75 ms)
[-----] 1 test from writeSortedRunsMultiple (75 ms total)

[-----] 1 test from sortRun
[ RUN ]   sortRun.test3
[ OK ]    sortRun.test3 (1 ms)
[-----] 1 test from sortRun (1 ms total)

[-----] 1 test from checkPriorityQueue
[ RUN ]   checkPriorityQueue.test4
[ OK ]    checkPriorityQueue.test4 (0 ms)
[-----] 1 test from checkPriorityQueue (0 ms total)

[-----] Global test environment tear-down
[*****] 4 tests from 4 test cases ran. (77 ms total)
[ PASSED ] 4 tests.
gloria@gloria-Inspiron-5593:/media/gloria/Work/UniversityOfFlorida/Study/SenII/DBI/Projects/Project2/Part1/a2test$
```