

# External Sorting Implemented

## Contributors:

*Megha Nagarmunoli UFID: 6768-1178*

*Siddharth Jain UFID: 7185-1161*

Steps to execute:

1. Unzip the file **SiddharthJain\_MeghaNagarmunoli\_p2\_1.zip**

2. cd **SiddharthJain\_MeghaNagarmunoli\_p2\_1/**

Note: **SiddharthJain\_MeghaNagarmunoli\_p2\_1** and **bin** should be in the same folder/level. **bin** folder should have all the **.bin** files required to execute the sort functionality.

3. To run the gTests,

1. make gtests.out

2. ./gtest.out

The following are the results of the gtests.

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

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

NumAtts =      1
  0:      0 Int
Running Test 3
./testdata/test.bin
Running Test 4
Number of errors : 0
[      OK ] test_sorting_for_a_run.testsorting (2 ms)
[-----] 1 test from test_sorting_for_a_run (2 ms total)

[-----] 1 test from test_sorting_for_a_run_1
[ RUN      ] test_sorting_for_a_run_1.testingsort
Running Test 1
./testdata/nation.bin

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

NumAtts =      1
  0:      2 Int
Running Test 3
./testdata/test.bin
Running Test 4
Number of errors : 0
[      OK ] test_sorting_for_a_run_1.testingsort (1 ms)
[-----] 1 test from test_sorting_for_a_run_1 (1 ms total)

[-----] 1 test from test_integration1
[ RUN      ] test_integration1.intergration_test

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

NumAtts =      1
  0:      2 Int
./testdata/nation.bin
producer: opened DBFile ./testdata/nation.bin
producer: inserted 25 recs into the pipe
Number of records25
Setting Indexes
Begin Merge
[      OK ] test_integration1.intergration_test (0 ms)
[-----] 1 test from test_integration1 (0 ms total)

[-----] 1 test from test_integration2
[ RUN      ] test_integration2.intergration_test

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

NumAtts =      1
  0:      1 Int
./testdata/lineitem.bin
producer: opened DBFile ./testdata/lineitem.bin
producer: inserted 60175 recs into the pipe
Number of records1660
Setting Indexes
Begin Merge
[      OK ] test_integration2.intergration_test (158 ms)
[-----] 1 test from test_integration2 (158 ms total)

[-----] Global test environment tear-down
[=====] 4 tests from 4 test cases ran. (161 ms total)
[ PASSED  ] 4 tests.
meghan@meghan-Inspiron-5584:~/Documents/gitWorkspace/database-system-implementation-2/a2test$
```

## Results of execution of ./runTestCases.sh

### Screenshot of output1.txt

```
output1.txt
1  n_nationkey: [19], n_name: [ROMANIA], n_regionkey: [3], n_comment: [ular asymptotes are about the furious multipliers. express dependencies nag above the i
2  n_nationkey: [22], n_name: [RUSSIA], n_regionkey: [3], n_comment: [ requests against the platelets use never according to the quickly regular pint]
3  n_nationkey: [23], n_name: [UNITED KINGDOM], n_regionkey: [3], n_comment: [eans boost carefully special requests. accounts are. carefull]
4  n_nationkey: [13], n_name: [JORDAN], n_regionkey: [4], n_comment: [ic deposits are blithely about the carefully regular pa]
5  n_nationkey: [4], n_name: [EGYPT], n_regionkey: [4], n_comment: [y above the carefully unusual theodolites. final dugouts are quickly across the furiously
6  n_nationkey: [11], n_name: [IRAQ], n_regionkey: [4], n_comment: [nic deposits boost atop the quickly final requests? quickly regula]
7  n_nationkey: [20], n_name: [SAUDI ARABIA], n_regionkey: [4], n_comment: [ts. silent requests haggle. closely express packages sleep across the blithely]
8  n_nationkey: [10], n_name: [IRAN], n_regionkey: [4], n_comment: [efully alongside of the slyly final dependencies. ]
9  consumer: removed 25 recs from the pipe
10 consumer: 25 recs out of 25 recs in sorted order
11 *****
12 consumer: removed 150000 recs from the pipe
13 Page count :234
14 consumer: 150000 recs out of 150000 recs in sorted order
15 *****
16 o_orderkey: [2185667], o_custkey: [51796], o_orderstatus: [F], o_totalprice: [511360], o_orderdate: [1992-10-08], o_orderpriority: [1-URGENT], o_clerk: [Cl
17 o_orderkey: [2199712], o_custkey: [66790], o_orderstatus: [0], o_totalprice: [515532], o_orderdate: [1996-09-30], o_orderpriority: [2-HIGH], o_clerk: [Cler
18 o_orderkey: [3586919], o_custkey: [24049], o_orderstatus: [F], o_totalprice: [522644], o_orderdate: [1992-11-07], o_orderpriority: [1-URGENT], o_clerk: [Cl
19 o_orderkey: [2232932], o_custkey: [13940], o_orderstatus: [0], o_totalprice: [522721], o_orderdate: [1997-04-13], o_orderpriority: [2-HIGH], o_clerk: [Cler
20 o_orderkey: [4576548], o_custkey: [108931], o_orderstatus: [0], o_totalprice: [525591], o_orderdate: [1997-12-26], o_orderpriority: [1-URGENT], o_clerk: [C
21 o_orderkey: [3043270], o_custkey: [144617], o_orderstatus: [0], o_totalprice: [530604], o_orderdate: [1997-02-12], o_orderpriority: [5-LOW], o_clerk: [Cler
22 o_orderkey: [4722021], o_custkey: [128120], o_orderstatus: [F], o_totalprice: [544089], o_orderdate: [1994-04-07], o_orderpriority: [1-URGENT], o_clerk: [C
23 o_orderkey: [1750466], o_custkey: [21433], o_orderstatus: [F], o_totalprice: [555285], o_orderdate: [1992-11-30], o_orderpriority: [4-NOT SPECIFIED], o_cle
24 consumer: removed 1500000 recs from the pipe
25 consumer: 1500000 recs out of 1500000 recs in sorted order
26 *****
27 n_nationkey: [19], n_name: [ROMANIA], n_regionkey: [3], n_comment: [ular asymptotes are about the furious multipliers. express dependencies nag above the i
28 n_nationkey: [22], n_name: [RUSSIA], n_regionkey: [3], n_comment: [ requests against the platelets use never according to the quickly regular pint]
29 n_nationkey: [23], n_name: [UNITED KINGDOM], n_regionkey: [3], n_comment: [eans boost carefully special requests. accounts are. carefull]
30 n_nationkey: [4], n_name: [EGYPT], n_regionkey: [4], n_comment: [y above the carefully unusual theodolites. final dugouts are quickly across the furiously
31 n_nationkey: [10], n_name: [IRAN], n_regionkey: [4], n_comment: [efully alongside of the slyly final dependencies. ]
32 n_nationkey: [11], n_name: [IRAQ], n_regionkey: [4], n_comment: [nic deposits boost atop the quickly final requests? quickly regula]
33 n_nationkey: [13], n_name: [JORDAN], n_regionkey: [4], n_comment: [ic deposits are blithely about the carefully regular pa]
34 n_nationkey: [20], n_name: [SAUDI ARABIA], n_regionkey: [4], n_comment: [ts. silent requests haggle. closely express packages sleep across the blithely]
35 consumer: removed 25 recs from the pipe
36 consumer: 25 recs out of 25 recs in sorted order
37
```

## Code Documentation

The BigQ constructor accepts the input pipe **in** and the output pipe **out**. It then creates another worker thread by the name **worker**. The worker thread executes a method described below

```
void* BigQ::SortAndMerge(void* arg) {
```

This method works in two phases.

- Phase 1- The following method executes phase 1

```
void BigQ::createFileWithRuns(Pipe* inputPipe, vector<pair<int,int>> &runMetadata, int &numRuns, int *runLength, OrderMaker *sort_order, File &new_file)
```

- The code reads the input pipe and extracts the records from it.
- Put the records in a vector until the vector has **runLength** worth of data.
- Then it sorts the data using the **orderMaker**.
- Use the sorted vector of records, it creates the pages from it i.e It creates a run and appends the run in the file. The following methods do this job

```
static void sortRun(vector<Record*> &records, File& new_file, int& gp_index, OrderMaker *sort_order)
```

- It repeatedly creates run and appends to the file along with maintaining the metadata about the start and end of each run which would be used to merge all the runs.
- Phase 2 - This phase uses the metadata about each run and merges the data using a priority queue.
  - The first record from the first page from the head of each run are put in a priority queue.
  - Records are extracted from the priority queue and its run is identified and the next record from that same run is put in the priority queue.
  - The extracted record from priority queue are put in the **outPipe** for the consumer to consume it and write it to another file. This marks the end of phase 2.