

proj2 Sorted File Implementation

- Haocheng Song(UFID: 11851321)
- Jiajing Liao(UFID: 01469951)

How to Run

For test.cc

```
make clean
make
./runTestCases.sh
```

For gtest

```
make clean
make
./gtest.out
```

major works explanation:

create new Architecture base on interface DBFileGeneric

DBFile is an unified class, it will determine the file type on Create() or Open(), there is also a .meta file as metadata

DBFile will be a entrance, it will help to find the real implementation class

DBFileGeneric is an interface, it only includes pure virtual functions

DBFileHeap, DBFileSorted, DBFileTree is the real implementation of DBFile, they contains real logics.

lead into new metadata, including filetype, runLength and orderMaker

int DBFile::Create (char *f_path, fType f_type, void *startup): convert orderMaker into metadata file

int DBFile::Open (char *f_path): convert a metadata file into orderMaker

Definition to metadata:

first row: file type

second row: runLen

rest of rows: orderMaker

Example metadata file:

```
sorted  # file type
16      # runLen
1       # orderMaker.numAtts
2       # orderMaker.whichAtts[0]
Int     # orderMaker.whichTypes[0]
```

we use a fstream to write out or parse in metadata

implement Sorted File

void DBFileSorted::writeMode(): inin the BigQ Thread

void DBFileSorted::readMode(): merge 2 sorted files into the new sorted file

As you can see in below Logic picture.

When a new record come in, we add it into the BigQ, and then BigQ will insert into into Runs and then get a sorted dif file. When it is necessary, we will merge the dif file with the original sorted file with a 2-Way merge algorithm and get the new sorted file.

However, we will not merge the dif file with sorted file instantly. Since merging is a time consuming processes, we will only do mergeing when there is a read-write change, or file close() operations.

Read-Write Switch detail

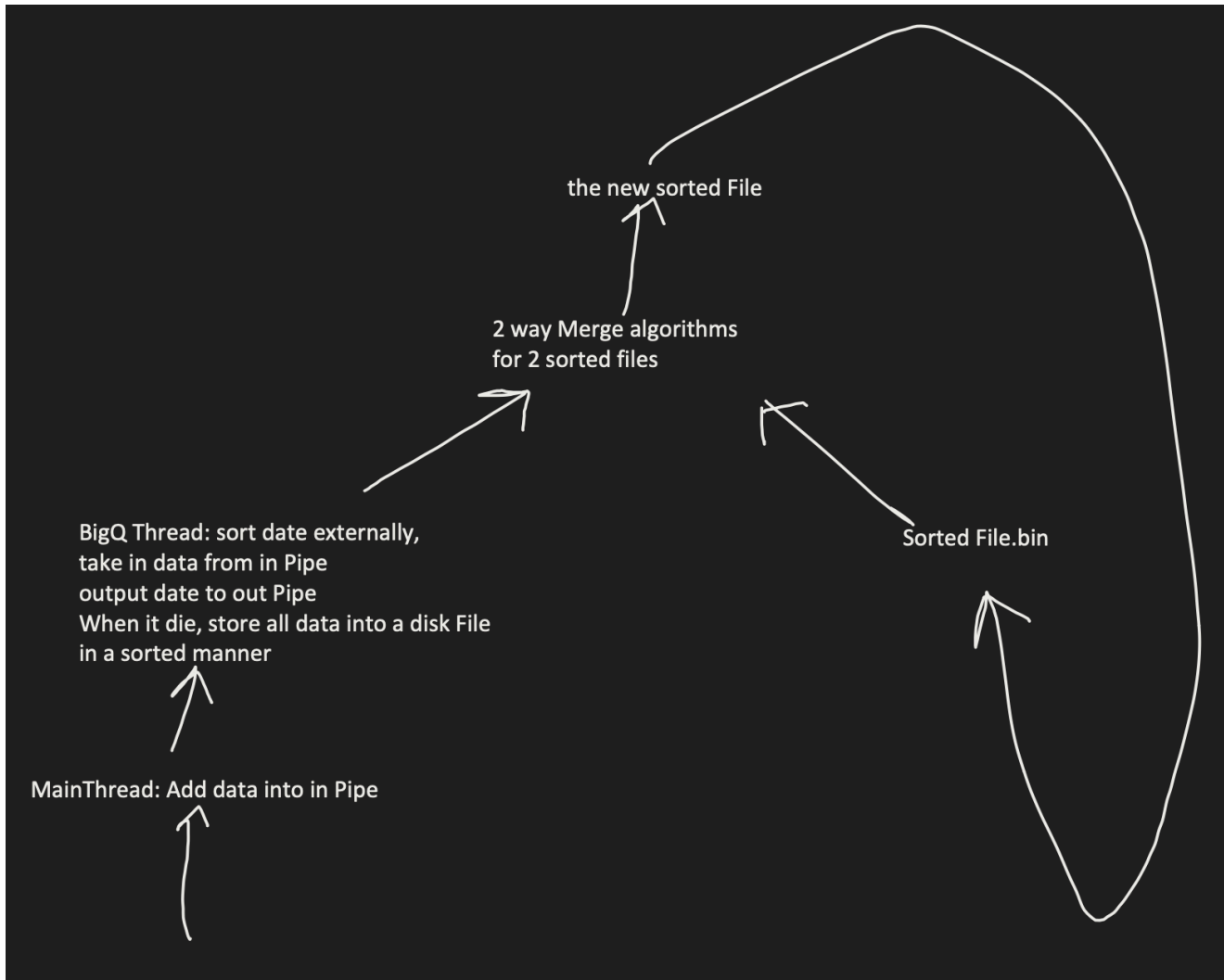
From read to write, it will set up the BigQ class in a new Thread for processing the incoming added Records

From write to read, it will kill the BigQ threads, merge the difFile with sorted File as the new sorted File.

File Merge detail

There are some temporarily files.

1. difFile
 2. mergeFile
 3. sortedFile
- I merge difFile and sortedFile into mergeFile using a 2-way merge algorithm, get the smaller one each time
 - Then I detele other files except the mergeFile, then I rename it into the right name.



implement of Binary Search

int DBFileSorted::GetNext (Record &fetchme, CNF &cnf, Record &literal):

I main 2 variables: lowerBound and higherBound, which are the possible minimum and maximum pageIndex

We scan all attributes in OrderMaker and CNF, find all the intersection attributes. Then for a LessThan operation, we find it's Upper Bound pageIndex. For a GreaterThan operation, we find it's Lower Bound pageIndex. As a result, the result must be in this range [lowerBound, higherBound], this will be a much smaller range compare to the original one. So, it will be much more efficient.

test on CISE computer

When compiling in CISE Machine, you have to delete `-o2`, otherwise there might be problems. My problem is that I cannot open my file again after I close it.

After below operations, everything is OK

gtest

we create 6 gtest cases and all of them passed.

Result picture

i) test1()

test.cc issues

as described here: https://ufl.instructure.com/courses/418634/discussion_topics/2913007

I find a problem in original code of project 2 part 2 code. Line 31 of test.cc (Links to an external site.), since it requires a CTRL+D, which is a "EOF", this make the consecutive "cin" not work. Attachment is the picture of endless while loop in Line 34 of test.cc (Links to an external site.)

As a result, I change sequence of test1(), move the CNF which requires a CTRL-D part as the last one to input, so the input.txt will be changed accordingly. test2() and test3() remain unchanged. The changed input.txt for test1() will be as below:

```
1
3
8
2
(c_phone)
```

```
2  IT IS IMPORTANT! MAKE SURE THE INFORMATION BELOW IS CORRECT!
3  catalog location:  catalog
4  tpch files dir:    ../tpch-dbgen/1G/
5  heap files dir:
6
7
8  select test option:
9      1. create sorted dbfile
10     2. scan a dbfile
11     3. run some query
12
13  select table:
14      1. nation
15      2. region
16      3. customer
```

```
17         4. part
18         5. partsupp
19         6. supplier
20         7. orders
21         8. lineitem
22
23     specify runlength:
24         input from file : ../tpch-dbgen/1G/customer.tbl
25
26     select option for : customer.bin
27         1. add a few (1 to 1k recs)
28         2. add a lot (1k to 1e+06 recs)
29         3. run some query
30
31     specify sort ordering (when done press ctrl-D):
32
33     output to dbfile : customer.bin
34     meta data in customer.bin.meta
35     NumAtts =      1
36     0:      4 String
37     begin writeMode
38     end writeMode
39     begin workerMain
40     1
41     meta data in tempDifFile.bin.meta
42     end workerMain
43     meta data in tempMergedFile.bin.meta
44
45         added 150000 recs..so far 150000
46
47     create finished.. 150000 recs inserted
48
```

ii) test2()

2

3

```
150031 c_custkey: [93747], c_name: [Customer#000093747], c_address: [u0M0NEzELIKCxbL219KerJmDpEjN2K9vXU], c_nationkey: [24], c_phone: [34-992-958-7084], c_acctbal: [9455.22], c_mktsegment: [FURNITURE], c_comment:
150032 c_custkey: [113698], c_name: [Customer#000113698], c_address: [I zTM8Gdh9aegsYsuQC6w4Sd7MXt8qff6 eR7], c_nationkey: [24], c_phone: [34-993-166-5163], c_acctbal: [-486.38], c_mktsegment: [HOUSEHOLD], c_comment:
150033 c_custkey: [1126905], c_name: [Customer#0001126905], c_address: [bGdhhyAUu, ], c_nationkey: [24], c_phone: [34-993-256-3681], c_acctbal: [7548.42], c_mktsegment: [MACHINERY], c_comment: [ly final instructions abo
150034 c_custkey: [29893], c_name: [Customer#000029893], c_address: [GGPC7fmgGqH4KhPrPnrSwy1lgR eV], c_nationkey: [24], c_phone: [34-993-311-3710], c_acctbal: [1260.4], c_mktsegment: [BUILDING], c_comment: [even pac
150035 c_custkey: [55993], c_name: [Customer#000055993], c_address: [AKSInlagfyceP0HlyE0c], c_nationkey: [24], c_phone: [34-993-429-1238], c_acctbal: [7623.36], c_mktsegment: [FURNITURE], c_comment: [ beans are silent
150036 c_custkey: [144536], c_name: [Customer#000144536], c_address: [4pt4pApv2Yn203K6dt895Kl], c_nationkey: [24], c_phone: [34-993-667-3229], c_acctbal: [5639.33], c_mktsegment: [HOUSEHOLD], c_comment: [s cajole care
150037 c_custkey: [46211], c_name: [Customer#000046211], c_address: [gH5iZB0ICv,9], c_nationkey: [24], c_phone: [34-993-926-3853], c_acctbal: [6337.97], c_mktsegment: [MACHINERY], c_comment: [l. fluffily special patte
150038 c_custkey: [43757], c_name: [Customer#000043757], c_address: [51ef5SR0XEYXlreJAK8], c_nationkey: [24], c_phone: [34-994-145-7928], c_acctbal: [-929.94], c_mktsegment: [MACHINERY], c_comment: [ ironic frets wake
150039 c_custkey: [113081], c_name: [Customer#000113081], c_address: [IE0y70MSYVrzzj] HWkL0p72Dc4ZB0M], c_nationkey: [24], c_phone: [34-994-432-9241], c_acctbal: [5935.36], c_mktsegment: [MACHINERY], c_comment: [lcky
150040 c_custkey: [95839], c_name: [Customer#000095839], c_address: [SvnZ1xk09h3081 98zNvLLctrS002CvZ], c_nationkey: [24], c_phone: [34-994-696-2809], c_acctbal: [1871.86], c_mktsegment: [BUILDING], c_comment: [n
150041 c_custkey: [126949], c_name: [Customer#000126949], c_address: [hH1qP9dNuYeYpITxkt], c_nationkey: [24], c_phone: [34-994-700-9912], c_acctbal: [8610.56], c_mktsegment: [BUILDING], c_comment: [slyly special fore
150042 c_custkey: [80619], c_name: [Customer#000080619], c_address: [y5kNm]8TNtWz8Ntpw 0RE], c_nationkey: [24], c_phone: [34-994-893-6724], c_acctbal: [70.76], c_mktsegment: [FURNITURE], c_comment: [tealthy packages
150043 c_custkey: [19768], c_name: [Customer#000019768], c_address: [RG5jIRgUUpITPHK5K302vcZpJ0X, ], c_nationkey: [24], c_phone: [34-995-194-4009], c_acctbal: [5206.51], c_mktsegment: [AUTOMOBILE], c_comment: [nsual r
150044 c_custkey: [57878], c_name: [Customer#000057878], c_address: [PaUN9uAmImntx70ha], c_nationkey: [24], c_phone: [34-995-249-6002], c_acctbal: [3861.85], c_mktsegment: [FURNITURE], c_comment: [al epitaphs detect
150045 c_custkey: [66195], c_name: [Customer#000066195], c_address: [YgacYJLreSUSIdygu1E348uAnxXayLTuS], c_nationkey: [24], c_phone: [34-995-870-2299], c_acctbal: [5186.45], c_mktsegment: [FURNITURE], c_comment: [f
150046 c_custkey: [74263], c_name: [Customer#000074263], c_address: [8BKuJXq4Ab1nXG0nJ,So6t6 ,ommxsDyGY0Jf], c_nationkey: [24], c_phone: [34-995-950-2170], c_acctbal: [7883.79], c_mktsegment: [MACHINERY], c_comment:
150047 c_custkey: [88858], c_name: [Customer#000088858], c_address: [IMshbshop], c_nationkey: [24], c_phone: [34-996-685-1799], c_acctbal: [6866.8], c_mktsegment: [AUTOMOBILE], c_comment: [theadolites after the fluff
150048 c_custkey: [139443], c_name: [Customer#000139443], c_address: [20M9KUIBcs,5P lTunh3V30,byvzJz0H0M], c_nationkey: [24], c_phone: [34-996-663-4670], c_acctbal: [-395.28], c_mktsegment: [BUILDING], c_comment: [
150049 c_custkey: [37226], c_name: [Customer#000037226], c_address: [D0K3oc1C7H29v4dKtHohh0KpJRUSS], c_nationkey: [24], c_phone: [34-996-695-7639], c_acctbal: [3155.24], c_mktsegment: [HOUSEHOLD], c_comment: [linal a
150050 c_custkey: [25985], c_name: [Customer#000025985], c_address: [m wL00 AZTfAYbRct0Z0AbJ], c_nationkey: [24], c_phone: [34-996-827-5819], c_acctbal: [6393.85], c_mktsegment: [BUILDING], c_comment: [sly across t
150051 c_custkey: [6048], c_name: [Customer#00006048], c_address: [DNHLKddov3MfAGa04I6c6Y1rS] NF], c_nationkey: [24], c_phone: [34-996-904-3634], c_acctbal: [2634.46], c_mktsegment: [AUTOMOBILE], c_comment: [l slee
150052 c_custkey: [2068], c_name: [Customer#00002068], c_address: [q3h08NHhsJ2n], c_nationkey: [24], c_phone: [34-996-906-1652], c_acctbal: [-839.34], c_mktsegment: [HOUSEHOLD], c_comment: [ the closely express execu
150053 c_custkey: [101355], c_name: [Customer#000101355], c_address: [sJ5taX90LFW], c_nationkey: [24], c_phone: [34-997-188-9666], c_acctbal: [32.59], c_mktsegment: [FURNITURE], c_comment: [even sheaves wake quickly
150054 c_custkey: [37888], c_name: [Customer#000037888], c_address: [IZkDfo3fodNGYonS0MJsJ], c_nationkey: [24], c_phone: [34-997-204-5897], c_acctbal: [2484.13], c_mktsegment: [AUTOMOBILE], c_comment: [al theadolites
150055 c_custkey: [3174], c_name: [Customer#00003174], c_address: [9XGPKFLWZMU], c_nationkey: [24], c_phone: [34-997-244-3200], c_acctbal: [6117.29], c_mktsegment: [BUILDING], c_comment: [round the furiously express
150056 c_custkey: [86364], c_name: [Customer#000086364], c_address: [VMK6m E2za75V3uVn1Zn0pJc9], c_nationkey: [24], c_phone: [34-997-251-4190], c_acctbal: [2691.67], c_mktsegment: [AUTOMOBILE], c_comment: [l even fox
150057 c_custkey: [35206], c_name: [Customer#000035206], c_address: [P HMX0TaHb3SVs], c_nationkey: [24], c_phone: [34-997-517-4660], c_acctbal: [320.81], c_mktsegment: [AUTOMOBILE], c_comment: [und the permanent p
150058 c_custkey: [48253], c_name: [Customer#000048253], c_address: [gAVuA89FV0ERqg0z 510c9WYd79YX1TLR6], c_nationkey: [24], c_phone: [34-997-697-4316], c_acctbal: [4429.89], c_mktsegment: [AUTOMOBILE], c_comment:
150059 c_custkey: [31537], c_name: [Customer#000031537], c_address: [8ZFr5r70wx2aS], c_nationkey: [24], c_phone: [34-998-198-8709], c_acctbal: [-252.7], c_mktsegment: [HOUSEHOLD], c_comment: [ructions cajole ideas, s
150060 c_custkey: [117513], c_name: [Customer#000117513], c_address: [cuB2 TE8E0NnPhm0dv6EhX0K43AV], c_nationkey: [24], c_phone: [34-998-315-4817], c_acctbal: [2334.66], c_mktsegment: [AUTOMOBILE], c_comment: [c ide
150061 c_custkey: [53730], c_name: [Customer#000053730], c_address: [6fv7lW9Pp9i0ntCZzws1e2GX0zJ9A ], c_nationkey: [24], c_phone: [34-998-382-6510], c_acctbal: [4923.4], c_mktsegment: [MACHINERY], c_comment: [ silen
150062 c_custkey: [12022], c_name: [Customer#000012022], c_address: [PwDMFqztkW0o], c_nationkey: [24], c_phone: [34-998-408-4382], c_acctbal: [2877.36], c_mktsegment: [HOUSEHOLD], c_comment: [layers. regular Pinto be
150063 c_custkey: [125210], c_name: [Customer#000125210], c_address: [B0JLJag79TVtSdxgPeb63S0qKa0v], c_nationkey: [24], c_phone: [34-998-711-2597], c_acctbal: [6678.32], c_mktsegment: [FURNITURE], c_comment: [ly ironi
150064 c_custkey: [50667], c_name: [Customer#000050667], c_address: [9aGwxCYLG6Gy2eA], c_nationkey: [24], c_phone: [34-999-115-2195], c_acctbal: [8764.52], c_mktsegment: [FURNITURE], c_comment: [mptotes around the sly
150065 c_custkey: [7288], c_name: [Customer#00007288], c_address: [JdZCqLe94Sq0jXAG7MXwRYyFWG54pFIRAshA], c_nationkey: [24], c_phone: [34-999-121-7718], c_acctbal: [82.07], c_mktsegment: [HOUSEHOLD], c_comment: [ns
150066 c_custkey: [24642], c_name: [Customer#000024642], c_address: [D4errr0HUVK], c_nationkey: [24], c_phone: [34-999-195-7029], c_acctbal: [6932.63], c_mktsegment: [BUILDING], c_comment: [notes above the pending, r
150067 c_custkey: [148611], c_name: [Customer#000148611], c_address: [H0NMKAELdZWMYQmQz2Z9g0ab7f0dNH54], c_nationkey: [24], c_phone: [34-999-283-6448], c_acctbal: [24.12], c_mktsegment: [HOUSEHOLD], c_comment: [re
150068 c_custkey: [102689], c_name: [Customer#000102689], c_address: [uPFKvGS0a20IL], c_nationkey: [24], c_phone: [34-999-363-7145], c_acctbal: [1626.07], c_mktsegment: [AUTOMOBILE], c_comment: [ly regular asymptotes
150069 c_custkey: [35478], c_name: [Customer#000035478], c_address: [0T5czyiLhZXXLvX4YZFeTo0H,lov6zYA], c_nationkey: [24], c_phone: [34-999-379-6816], c_acctbal: [6097.18], c_mktsegment: [MACHINERY], c_comment: [it
150070 c_custkey: [103505], c_name: [Customer#000103505], c_address: [XPMXpdCD7HRqcV8ABpJF3ix0c9A6X ZB2 ], c_nationkey: [24], c_phone: [34-999-404-9770], c_acctbal: [582.88], c_mktsegment: [FURNITURE], c_comment: [ful
150071 c_custkey: [62787], c_name: [Customer#000062787], c_address: [bRbtSTIXhh], c_nationkey: [24], c_phone: [34-999-473-9448], c_acctbal: [7429.32], c_mktsegment: [FURNITURE], c_comment: [zzle slyly after the slyly
150072 c_custkey: [41774], c_name: [Customer#000041774], c_address: [DpHfJFKVkr4LQelf9T5SSKzD0u4G3asKeya5M7d2], c_nationkey: [24], c_phone: [34-999-535-1203], c_acctbal: [7118.29], c_mktsegment: [FURNITURE], c_comment
150073 c_custkey: [94388], c_name: [Customer#000094388], c_address: [5aE9rPeLngevGv8KcGy0r,NH0TSuC4], c_nationkey: [24], c_phone: [34-999-618-6653], c_acctbal: [8636.13], c_mktsegment: [FURNITURE], c_comment: [g, eve
150074 c_custkey: [102414], c_name: [Customer#000102414], c_address: [Xz96iWbsxiW,LCUCYAcP0ENGLUQt], c_nationkey: [24], c_phone: [34-999-618-6881], c_acctbal: [9614.62], c_mktsegment: [MACHINERY], c_comment: [usly
150075
150076 scanned 150000 recs
```

iii) test3()

3

3

(c_phone > '34-999-195-7029') AND (c_mktsegment = 'FURNITURE')

```
150079 *****
150080
150081 ** IMPORTANT: MAKE SURE THE INFORMATION BELOW IS CORRECT **
150082 catalog location: catalog
150083 tpch files dir: ../tpch-dbggen/1G/
150084 heap files dir:
150085
150086
150087 select test option:
150088 1. create sorted dbfile
150089 2. scan a dbfile
150090 3. run some query
150091
150092 select table:
150093 1. nation
150094 2. region
150095 3. customer
150096 4. part
150097 5. partsupp
150098 6. supplier
150099 7. orders
150100 8. lineitem
150101
150102 enter CNF predicate (when done press ctrl-D):
150103 NumAtts = 1
150104 0: 4 String
150105 ( Att 4 from left record > Att 0 from literal record (String)) AND
150106 ( Att 6 from left record = Att 1 from literal record (String))
150107 original lower bounds 0
150108 original higher bounds: 234
150109 updated lower bound by Binary Search: 233
150110 updated higher bound by Binary Search: 234
150111 c_custkey: [103505], c_name: [Customer#000103505], c_address: [XPMXpdCD7HRqcV8ABpJF3ix0c9A6X ZB2 ], c_nationkey: [24], c_phone: [34-999-404-9770], c_acctbal: [582.88], c_mktsegment: [FURNITURE], c_comment: [ful
150112 c_custkey: [62787], c_name: [Customer#000062787], c_address: [bRbtSTIXhh], c_nationkey: [24], c_phone: [34-999-473-9448], c_acctbal: [7429.32], c_mktsegment: [FURNITURE], c_comment: [zzle slyly after the slyly
150113 c_custkey: [41774], c_name: [Customer#000041774], c_address: [DpHfJFKVkr4LQelf9T5SSKzD0u4G3asKeya5M7d2], c_nationkey: [24], c_phone: [34-999-535-1203], c_acctbal: [7118.29], c_mktsegment: [FURNITURE], c_comment
150114 c_custkey: [94388], c_name: [Customer#000094388], c_address: [5aE9rPeLngevGv8KcGy0r,NH0TSuC4], c_nationkey: [24], c_phone: [34-999-618-6653], c_acctbal: [8636.13], c_mktsegment: [FURNITURE], c_comment: [g, eve
150115
150116 query over customer.bin returned 4 recs
150117
150118
150119
```

d) gtest

```
(base) abcdeg:a2-2test ljj$ ./gtest.out
[=====] Running 6 tests from 1 test suite.
[-----] Global test environment set-up.
[-----] 6 tests from DBFileTest
[ RUN      ] DBFileTest.CreateTest
meta data in ./gtest.bin.meta
begin c Create
end Create
==1
[      OK ] DBFileTest.CreateTest (1 ms)
[ RUN      ] DBFileTest.OpenTest
NumAtts =      1
  0:      4 String
[      OK ] DBFileTest.OpenTest (1 ms)
[ RUN      ] DBFileTest.CloseTest
NumAtts =      1
  0:      4 String
[      OK ] DBFileTest.CloseTest (0 ms)
[ RUN      ] DBFileTest.OpenAndCloseTest
NumAtts =      1
  0:      4 String
[      OK ] DBFileTest.OpenAndCloseTest (1 ms)
[ RUN      ] DBFileTest.CreateAndCloseTest
meta data in ./gtest.bin.meta
begin c Create
end Create
[      OK ] DBFileTest.CreateAndCloseTest (0 ms)
[ RUN      ] DBFileTest.GetNextTest
NumAtts =      1
  0:      4 String
```

```
[      OK ] DBFileTest.GetNextTest (0 ms)
[-----] 6 tests from DBFileTest (3 ms total)

[-----] Global test environment tear-down
[=====] 6 tests from 1 test suite ran. (3 ms total)
[  PASSED  ] 6 tests.
```