RocksDB Festival

RocksDB Env

Supported by IITP, StarLab.

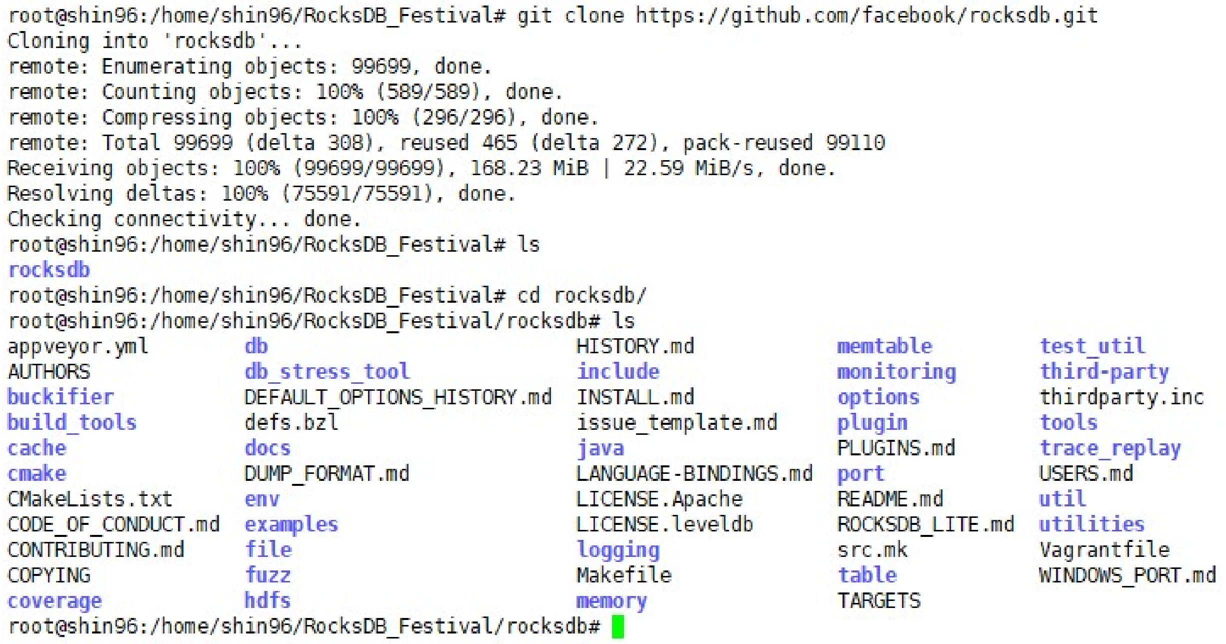
July 5, 2021

Hojin Shin, Jongmoo Choi
choijm@dankook.ac.kr
http://embedded.dankook.ac.kr/~choijm

RocksDB Festival

Git clone RocksDB

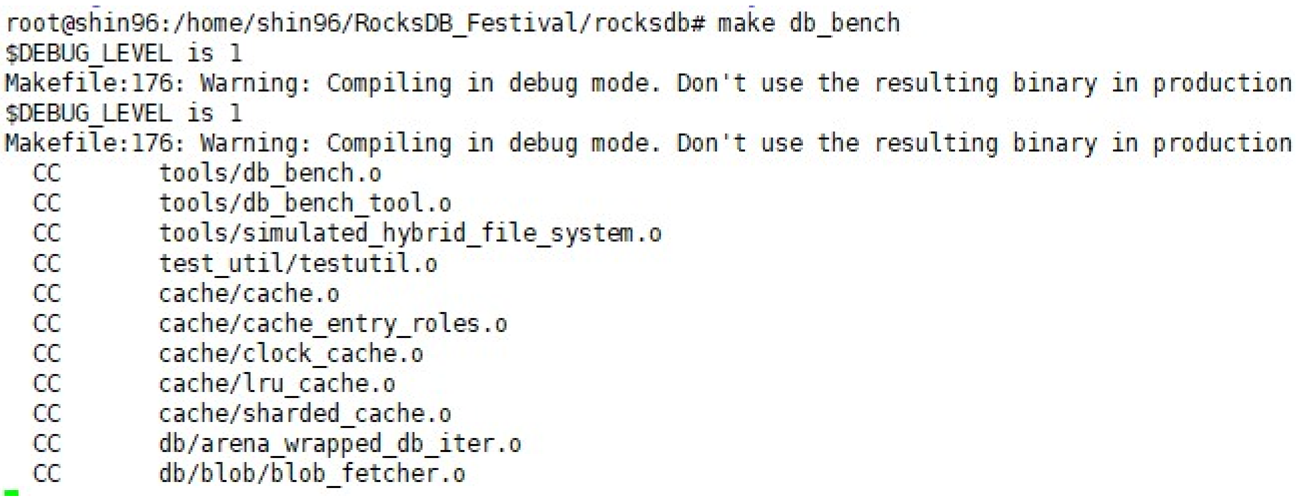
- ✓ git clone https://github.com/facebook/rocksdb.git
- ✓ cd rocksdb



```
root@shin96:/home/shin96/RocksDB_Festival# git clone https://github.com/facebook/rocksdb.git
Cloning into 'rocksdb'...
remote: Enumerating objects: 99699, done.
remote: Counting objects: 100% (589/589), done.
remote: Compressing objects: 100% (296/296), done.
remote: Total 99699 (delta 308), reused 465 (delta 272), pack-reused 99110
Receiving objects: 100% (99699/99699), 168.23 MiB | 22.59 MiB/s, done.
Resolving deltas: 100% (75591/75591), done.
Checking connectivity... done.
root@shin96:/home/shin96/RocksDB_Festival# ls
rocksdb
root@shin96:/home/shin96/RocksDB_Festival# cd rocksdb/
root@shin96:/home/shin96/RocksDB_Festival/rocksdb# ls
appveyor.yml        db                          HISTORY.md            memtable         test_util
AUTHORS             db_stress_tool              include               monitoring       third-party
buckifier           DEFAULT_OPTIONS_HISTORY.md  INSTALL.md            options          thirdparty.inc
build_tools         defs.bzl                    issue_template.md     plugin           tools
cache               docs                        java                  PLUGINS.md       trace_replay
cmake               DUMP_FORMAT.md              LANGUAGE-BINDINGS.md  port             USERS.md
CMakeLists.txt      env                         LICENSE.Apache        README.md        util
CODE_OF_CONDUCT.md  examples                    LICENSE.leveldb       ROCKSDB_LITE.md  utilities
CONTRIBUTING.md     file                        logging               src.mk           Vagrantfile
COPYING             fuzz                        Makefile              table            WINDOWS_PORT.md
coverage            hdfs                        memory                TARGETS
root@shin96:/home/shin96/RocksDB_Festival/rocksdb#
```

RocksDB Festival

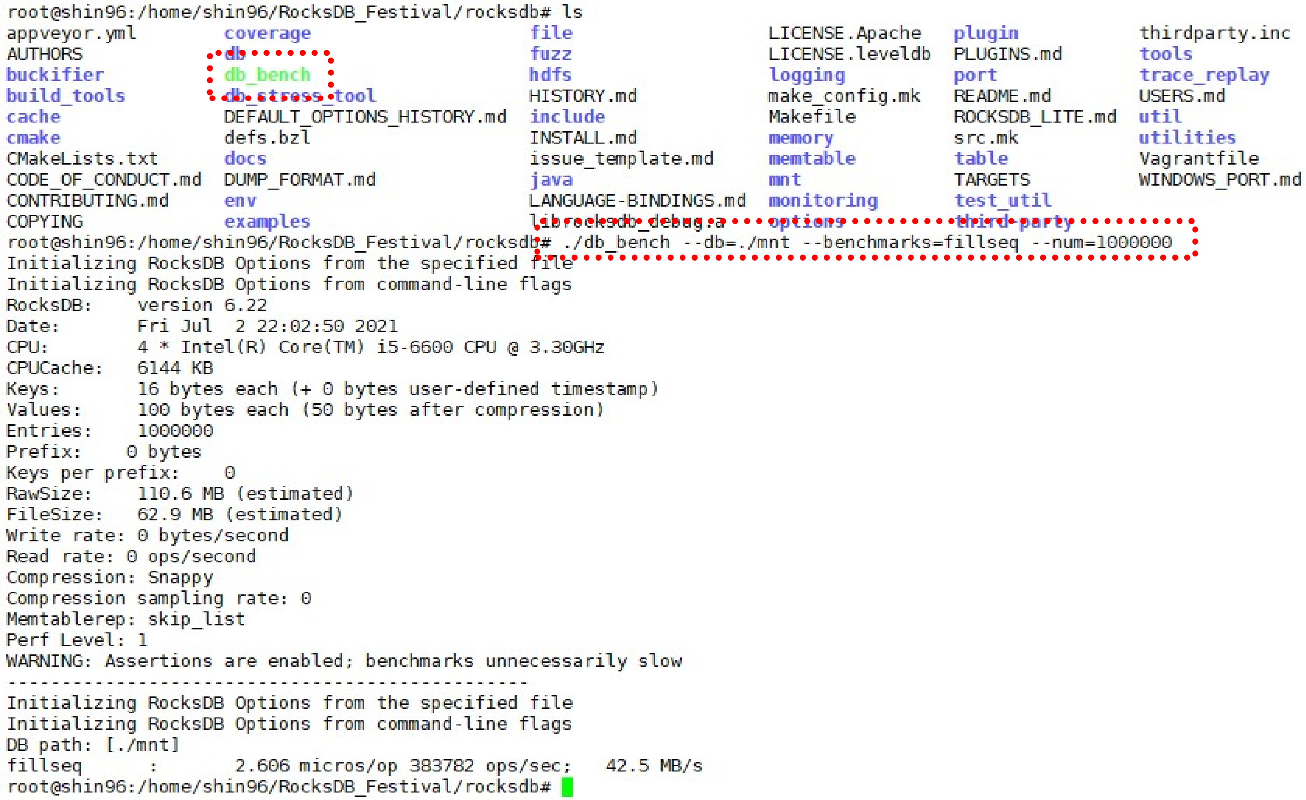
- Git clone RocksDB cont’
  - ✓ make db_bench
  - ✓ Then you can use RocksDB db_bench
  - ✓ It takes long time



```
root@shin96:/home/shin96/RocksDB_Festival/rocksdb# make db_bench
$DEBUG_LEVEL is 1
Makefile:176: Warning: Compiling in debug mode. Don't use the resulting binary in production
$DEBUG_LEVEL is 1
Makefile:176: Warning: Compiling in debug mode. Don't use the resulting binary in production
  CC      tools/db_bench.o
  CC      tools/db_bench_tool.o
  CC      tools/simulated_hybrid_file_system.o
  CC      test_util/testutil.o
  CC      cache/cache.o
  CC      cache/cache_entry_roles.o
  CC      cache/clock_cache.o
  CC      cache/lru_cache.o
  CC      cache/sharded_cache.o
  CC      db/arena_wrapped_db_iter.o
  CC      db/blob/blob_fetcher.o
```

RocksDB Festival

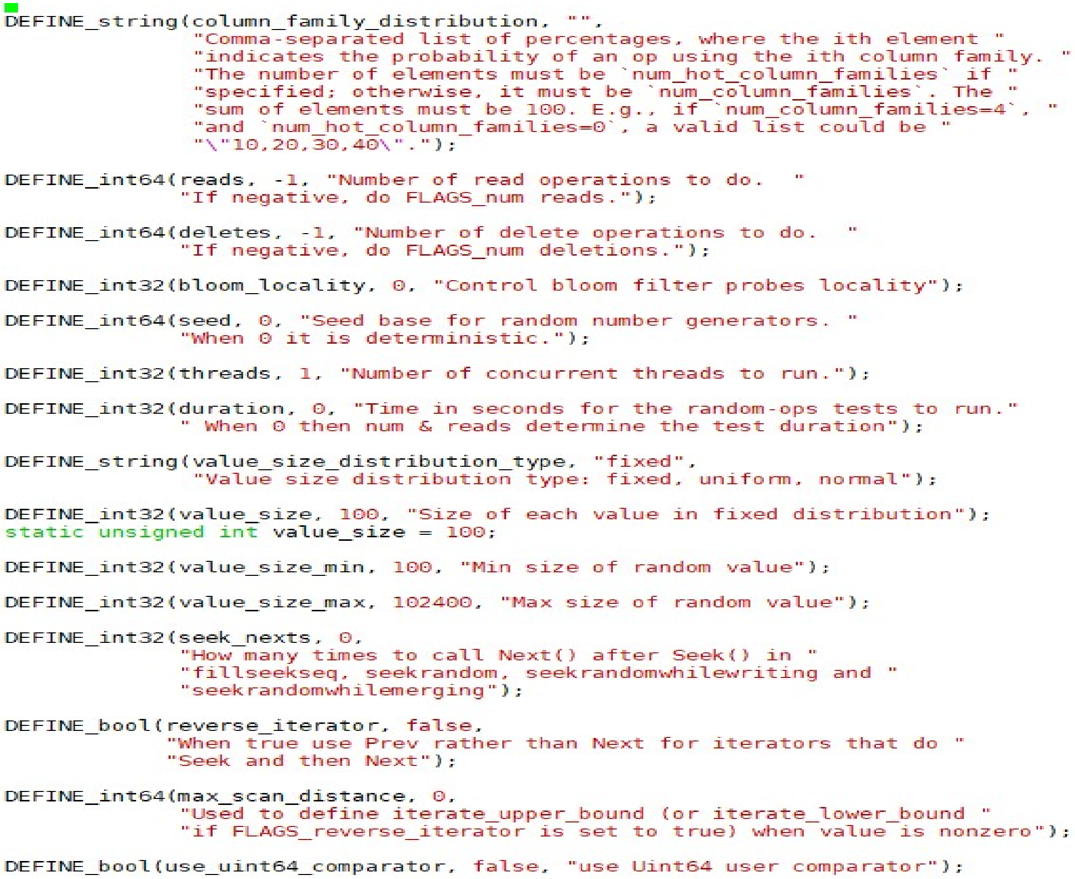
- Git clone RocksDB cont’
  - ✓ ./db_bench ~~something options
  - ✓ If you want set your device, you use [--db=(DEV_PATH)] option
  - ✓ Example)



```
root@shin96:/home/shin96/RocksDB_Festival/rocksdb# ls
appveyor.yml        coverage                    file                  LICENSE.Apache   plugin           thirdparty.inc
AUTHORS             db                          fuzz                  LICENSE.leveldb  PLUGINS.md       tools
buckifier           db_bench                    hdfs                  logging          port             trace_replay
build_tools         db_stress_tool              HISTORY.md            make_config.mk   README.md        USERS.md
cache               DEFAULT_OPTIONS_HISTORY.md  include               Makefile         ROCKSDB_LITE.md  util
cmake               defs.bzl                    INSTALL.md            memory           src.mk           utilities
CMakeLists.txt      docs                        issue_template.md     memtable         table            Vagrantfile
CODE_OF_CONDUCT.md  DUMP_FORMAT.md              java                  mnt              TARGETS          WINDOWS_PORT.md
CONTRIBUTING.md     env                         LANGUAGE-BINDINGS.md  monitoring       test_util
COPYING             examples                    librocksdb_debug.a    options          third-party
root@shin96:/home/shin96/RocksDB_Festival/rocksdb# ./db_bench --db=./mnt --benchmarks=fillseq --num=1000000
Initializing RocksDB Options from the specified file
Initializing RocksDB Options from command-line flags
RocksDB:    version 6.22
Date:       Fri Jul  2 22:02:50 2021
CPU:        4 * Intel(R) Core(TM) i5-6600 CPU @ 3.30GHz
CPUCache:   6144 KB
Keys:       16 bytes each (+ 0 bytes user-defined timestamp)
Values:     100 bytes each (50 bytes after compression)
Entries:    1000000
Prefix:    0 bytes
Keys per prefix:    0
RawSize:    110.6 MB (estimated)
FileSize:   62.9 MB (estimated)
Write rate: 0 bytes/second
Read rate: 0 ops/second
Compression: Snappy
Compression sampling rate: 0
Memtablerep: skip_list
Perf Level: 1
WARNING: Assertions are enabled; benchmarks unnecessarily slow
------------------------------------------------
Initializing RocksDB Options from the specified file
Initializing RocksDB Options from command-line flags
DB path: [./mnt]
fillseq      :       2.606 micros/op 383782 ops/sec;    42.5 MB/s
root@shin96:/home/shin96/RocksDB_Festival/rocksdb#
```

RocksDB Festival

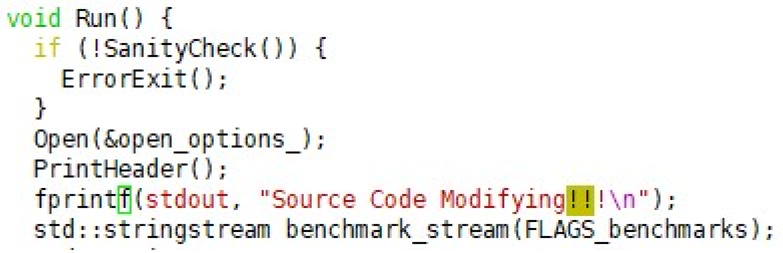
- Git clone RocksDB cont’
  - ✓ You can find db_bench options in tools/db_bench_tool.cc
  - ✓ Example)



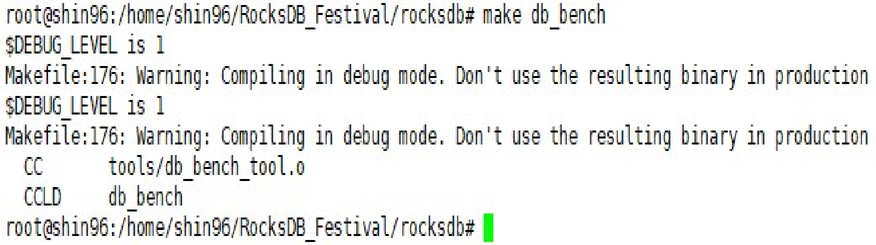
```
DEFINE_string(column_family_distribution, "",
              "Comma-separated list of percentages, where the ith element "
              "indicates the probability of an op using the ith column family. "
              "The number of elements must be `num_hot_column_families` if "
              "specified; otherwise, it must be `num_column_families`. The "
              "sum of elements must be 100. E.g., if `num_column_families=4`, "
              "and `num_hot_column_families=0`, a valid list could be "
              "\"10,20,30,40\".");

DEFINE_int64(reads, -1, "Number of read operations to do.  "
             "If negative, do FLAGS_num reads.");

DEFINE_int64(deletes, -1, "Number of delete operations to do.  "
             "If negative, do FLAGS_num deletions.");

DEFINE_int32(bloom_locality, 0, "Control bloom filter probes locality");

DEFINE_int64(seed, 0, "Seed base for random number generators. "
             "When 0 it is deterministic.");

DEFINE_int32(threads, 1, "Number of concurrent threads to run.");

DEFINE_int32(duration, 0, "Time in seconds for the random-ops tests to run."
             " When 0 then num & reads determine the test duration");

DEFINE_string(value_size_distribution_type, "fixed",
              "Value size distribution type: fixed, uniform, normal");

DEFINE_int32(value_size, 100, "Size of each value in fixed distribution");
static unsigned int value_size = 100;

DEFINE_int32(value_size_min, 100, "Min size of random value");

DEFINE_int32(value_size_max, 102400, "Max size of random value");

DEFINE_int32(seek_nexts, 0,
             "How many times to call Next() after Seek() in "
             "fillseekseq, seekrandom, seekrandomwhilewriting and "
             "seekrandomwhilemerging");

DEFINE_bool(reverse_iterator, false,
            "When true use Prev rather than Next for iterators that do "
            "Seek and then Next");

DEFINE_int64(max_scan_distance, 0,
             "Used to define iterate_upper_bound (or iterate_lower_bound "
             "if FLAGS_reverse_iterator is set to true) when value is nonzero");

DEFINE_bool(use_uint64_comparator, false, "use Uint64 user comparator");
```

RocksDB Festival

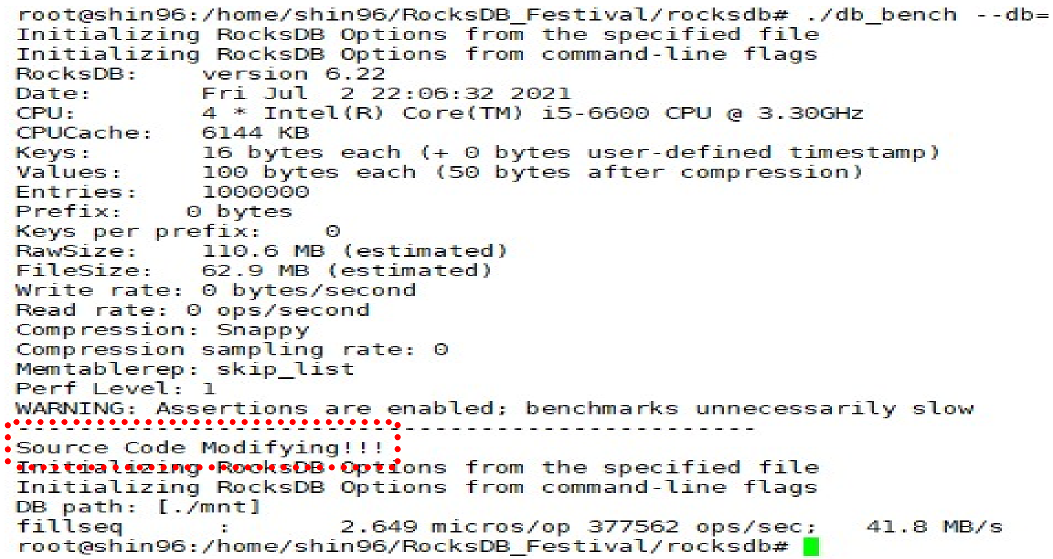
- Git clone RocksDB cont'
  - ✓ After modifying the source code, compilation must be performed again through make db_bench
  - ✓ Example)



```
void Run() {
  if (!SanityCheck()) {
    ErrorExit();
  }
  Open(&open_options_);
  PrintHeader();
  fprintf(stdout, "Source Code Modifying!!!\n");
  std::stringstream benchmark_stream(FLAGS_benchmarks);
```



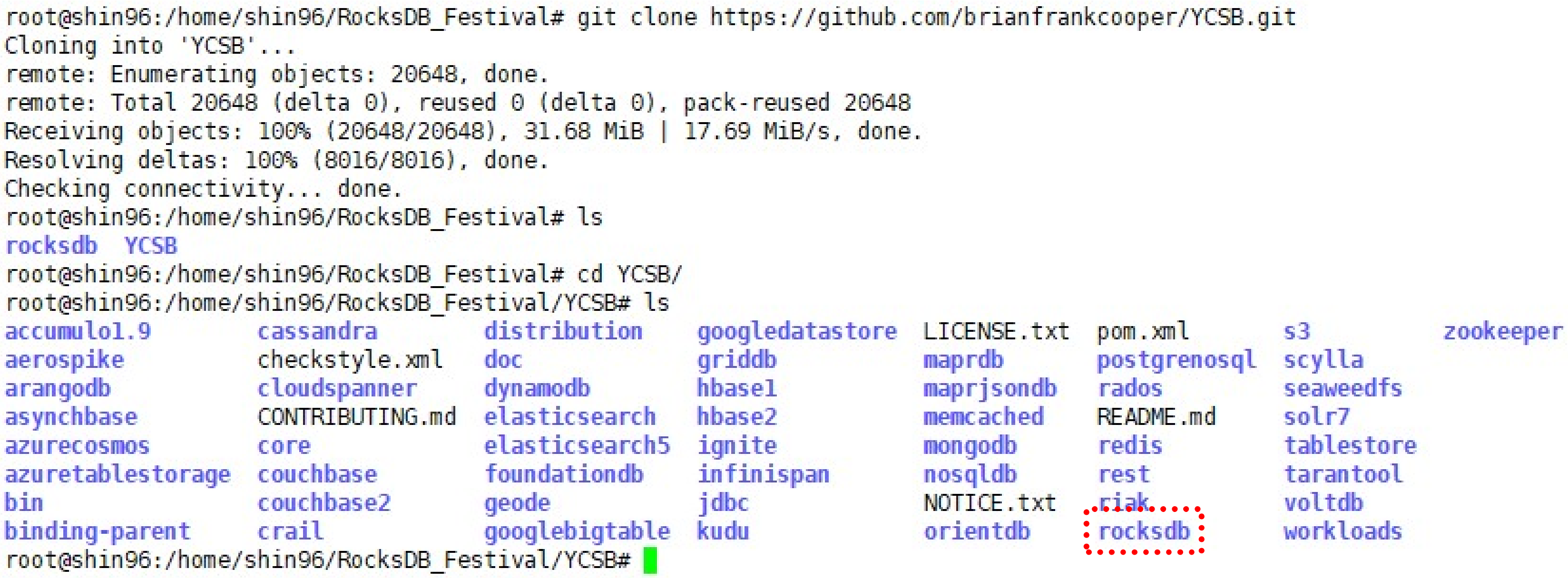
```
root@shin96:/home/shin96/RocksDB_Festival/rocksdb# make db_bench
$DEBUG_LEVEL is 1
Makefile:176: Warning: Compiling in debug mode. Don't use the resulting binary in production
$DEBUG_LEVEL is 1
Makefile:176: Warning: Compiling in debug mode. Don't use the resulting binary in production
  CC      tools/db_bench_tool.o
  CCLD    db_bench
root@shin96:/home/shin96/RocksDB_Festival/rocksdb#
```



```
root@shin96:/home/shin96/RocksDB_Festival/rocksdb# ./db_bench --db=
Initializing RocksDB Options from the specified file
Initializing RocksDB Options from command-line flags
RocksDB:    version 6.22
Date:       Fri Jul  2 22:06:32 2021
CPU:        4 * Intel(R) Core(TM) i5-6600 CPU @ 3.30GHz
CPUCache:   6144 KB
Keys:       16 bytes each (+ 0 bytes user-defined timestamp)
Values:     100 bytes each (50 bytes after compression)
Entries:    1000000
Prefix:    0 bytes
Keys per prefix:    0
RawSize:    110.6 MB (estimated)
FileSize:   62.9 MB (estimated)
Write rate: 0 bytes/second
Read rate: 0 ops/second
Compression: Snappy
Compression sampling rate: 0
Memtablerep: skip_list
Perf Level: 1
WARNING: Assertions are enabled; benchmarks unnecessarily slow
------------------------------------------------
Source Code Modifying!!!
Initializing RocksDB Options from the specified file
Initializing RocksDB Options from command-line flags
DB path: [./mnt]
fillseq      :       2.649 micros/op 377562 ops/sec;    41.8 MB/s
root@shin96:/home/shin96/RocksDB_Festival/rocksdb#
```

RocksDB Festival

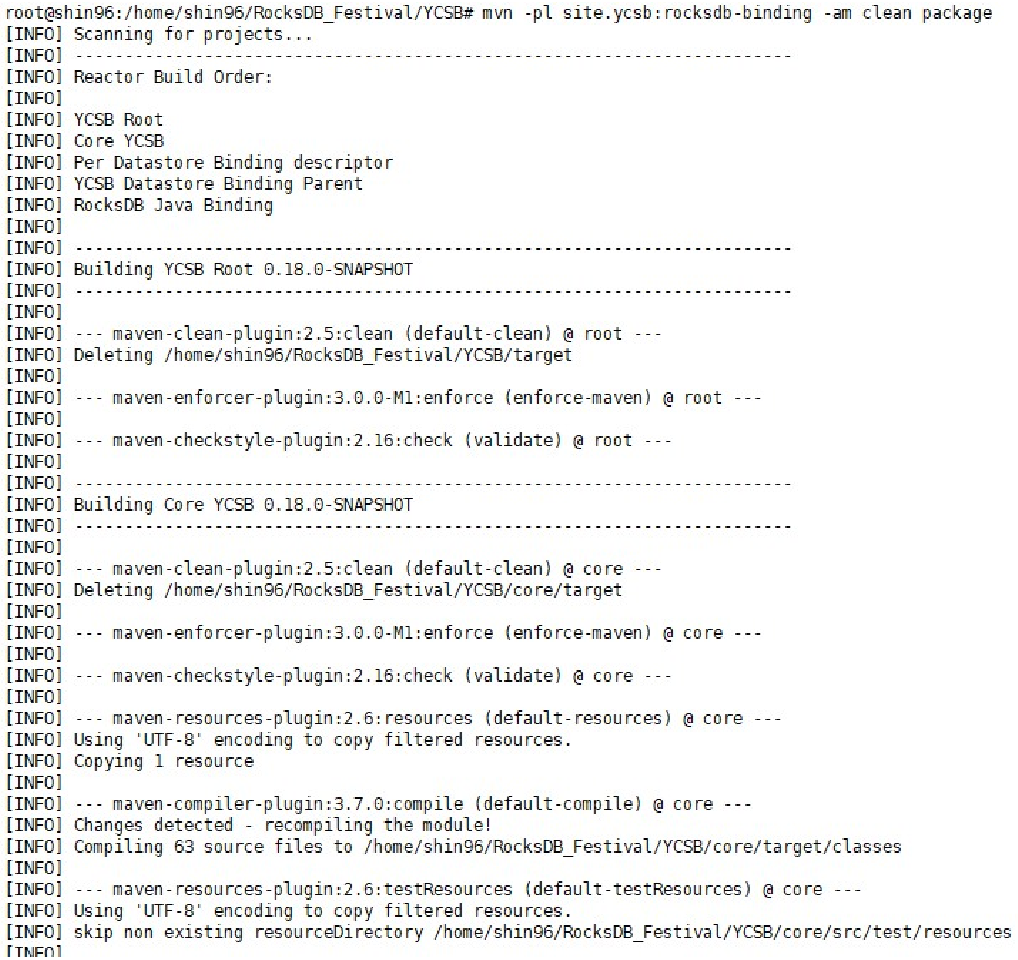
- YCSB
  - ✓ git clone https://github.com/brianfrankcooper/YCSB.git
  - ✓ cd YCSB



```
root@shin96:/home/shin96/RocksDB_Festival# git clone https://github.com/brianfrankcooper/YCSB.git
Cloning into 'YCSB'...
remote: Enumerating objects: 20648, done.
remote: Total 20648 (delta 0), reused 0 (delta 0), pack-reused 20648
Receiving objects: 100% (20648/20648), 31.68 MiB | 17.69 MiB/s, done.
Resolving deltas: 100% (8016/8016), done.
Checking connectivity... done.
root@shin96:/home/shin96/RocksDB_Festival# ls
rocksdb  YCSB
root@shin96:/home/shin96/RocksDB_Festival# cd YCSB/
root@shin96:/home/shin96/RocksDB_Festival/YCSB# ls
accumulo1.9        cassandra        distribution    googledatastore  LICENSE.txt  pom.xml       s3          zookeeper
aerospike          checkstyle.xml   doc             griddb           maprdb       postgrenosql  scylla
arangodb           cloudspanner     dynamodb        hbase1           maprjsondb   rados         seaweedfs
asynchbase         CONTRIBUTING.md  elasticsearch   hbase2           memcached    README.md     solr7
azurecosmos        core             elasticsearch5  ignite           mongodb      redis         tablestore
azuretablestorage  couchbase        foundationdb    infinispan       nosqldb      rest          tarantool
bin                couchbase2       geode           jdbc             NOTICE.txt   riak          voltdb
binding-parent     crail            googlebigtable  kudu             orientdb     rocksdb       workloads
root@shin96:/home/shin96/RocksDB_Festival/YCSB# 
```

RocksDB Festival

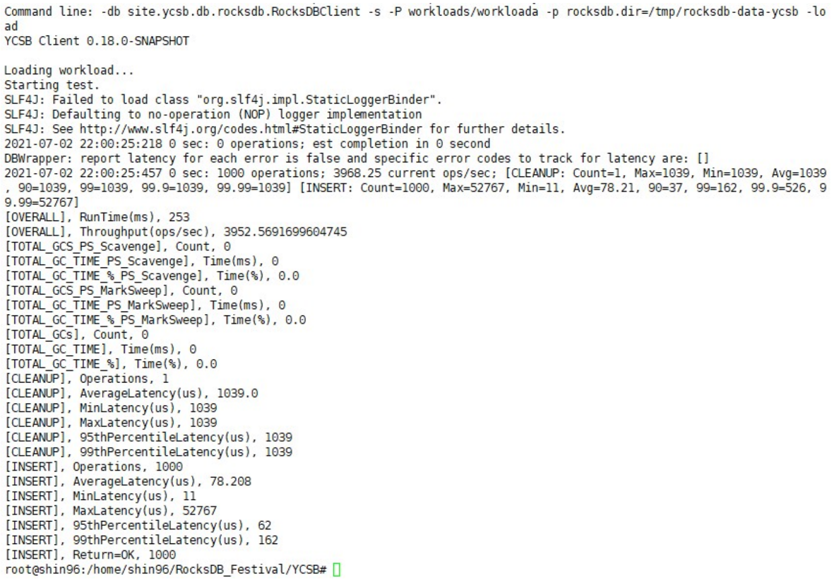
- YCSB cont’
  - ✓ mvn –pl sit.ycsb:rocksdb-binding –am clean package



```
root@shin96:/home/shin96/RocksDB_Festival/YCSB# mvn -pl site.ycsb:rocksdb-binding -am clean package
[INFO] Scanning for projects...
[INFO] ------------------------------------------------------------------------
[INFO] Reactor Build Order:
[INFO]
[INFO] YCSB Root
[INFO] Core YCSB
[INFO] Per Datastore Binding descriptor
[INFO] YCSB Datastore Binding Parent
[INFO] RocksDB Java Binding
[INFO]
[INFO] ------------------------------------------------------------------------
[INFO] Building YCSB Root 0.18.0-SNAPSHOT
[INFO] ------------------------------------------------------------------------
[INFO]
[INFO] --- maven-clean-plugin:2.5:clean (default-clean) @ root ---
[INFO] Deleting /home/shin96/RocksDB_Festival/YCSB/target
[INFO]
[INFO] --- maven-enforcer-plugin:3.0.0-M1:enforce (enforce-maven) @ root ---
[INFO]
[INFO] --- maven-checkstyle-plugin:2.16:check (validate) @ root ---
[INFO]
[INFO] ------------------------------------------------------------------------
[INFO] Building Core YCSB 0.18.0-SNAPSHOT
[INFO] ------------------------------------------------------------------------
[INFO]
[INFO] --- maven-clean-plugin:2.5:clean (default-clean) @ core ---
[INFO] Deleting /home/shin96/RocksDB_Festival/YCSB/core/target
[INFO]
[INFO] --- maven-enforcer-plugin:3.0.0-M1:enforce (enforce-maven) @ core ---
[INFO]
[INFO] --- maven-checkstyle-plugin:2.16:check (validate) @ core ---
[INFO]
[INFO] --- maven-resources-plugin:2.6:resources (default-resources) @ core ---
[INFO] Using 'UTF-8' encoding to copy filtered resources.
[INFO] Copying 1 resource
[INFO]
[INFO] --- maven-compiler-plugin:3.7.0:compile (default-compile) @ core ---
[INFO] Changes detected - recompiling the module!
[INFO] Compiling 63 source files to /home/shin96/RocksDB_Festival/YCSB/core/target/classes
[INFO]
[INFO] --- maven-resources-plugin:2.6:testResources (default-testResources) @ core ---
[INFO] Using 'UTF-8' encoding to copy filtered resources.
[INFO] skip non existing resourceDirectory /home/shin96/RocksDB_Festival/YCSB/core/src/test/resources
[INFO]
```

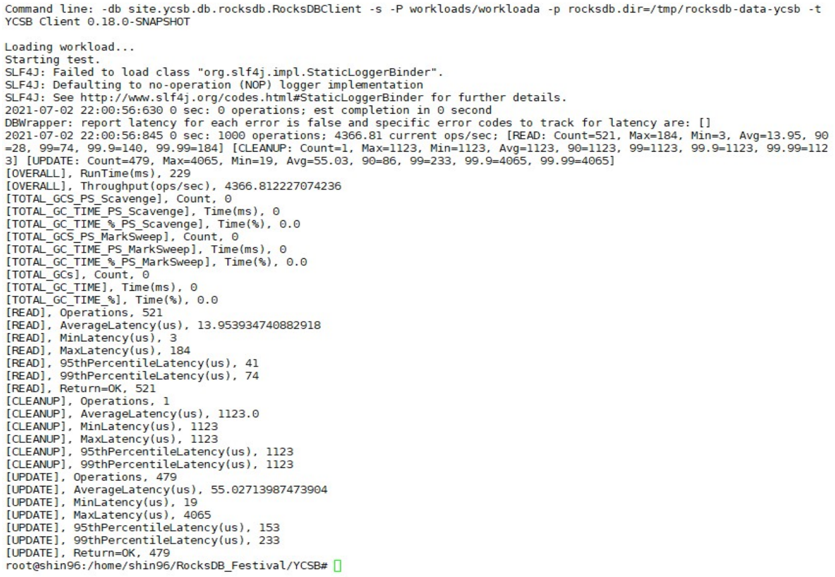
RocksDB Festival

- YCSB cont’
  - ✓ load workload
    - ./bin/ycsb load rocksdb –s –P workloads/workloada –p rocksdb.dir=[PATH] –p recordcount=[num]
  - ✓ run workload
    - ./bin/ycsb run rocksdb –s –P workloads/workloada –p rocksdb.dir=[PATH] –p operationcount=[num]



```
Command line: -db site.ycsb.db.rocksdb.RocksDBClient -s -P workloads/workloada -p rocksdb.dir=/tmp/rocksdb-data-ycsb -lo
ad
YCSB Client 0.18.0-SNAPSHOT

Loading workload...
Starting test.
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
2021-07-02 22:00:25:218 0 sec: 0 operations; est completion in 0 second
DBWrapper: report latency for each error is false and specific error codes to track for latency are: []
2021-07-02 22:00:25:457 0 sec: 1000 operations; 3968.25 current ops/sec; [CLEANUP: Count=1, Max=1039, Min=1039, Avg=1039
, 90=1039, 99=1039, 99.9=1039, 99.99=1039] [INSERT: Count=1000, Max=52767, Min=11, Avg=78.21, 90=37, 99=162, 99.9=526, 9
9.99=52767]
[OVERALL], RunTime(ms), 253
[OVERALL], Throughput(ops/sec), 3952.5691699604745
[TOTAL_GCS_PS_Scavenge], Count, 0
[TOTAL_GC_TIME_PS_Scavenge], Time(ms), 0
[TOTAL_GC_TIME_%_PS_Scavenge], Time(%), 0.0
[TOTAL_GCS_PS_MarkSweep], Count, 0
[TOTAL_GC_TIME_PS_MarkSweep], Time(ms), 0
[TOTAL_GC_TIME_%_PS_MarkSweep], Time(%), 0.0
[TOTAL_GCs], Count, 0
[TOTAL_GC_TIME], Time(ms), 0
[TOTAL_GC_TIME_%], Time(%), 0.0
[CLEANUP], Operations, 1
[CLEANUP], AverageLatency(us), 1039.0
[CLEANUP], MinLatency(us), 1039
[CLEANUP], MaxLatency(us), 1039
[CLEANUP], 95thPercentileLatency(us), 1039
[CLEANUP], 99thPercentileLatency(us), 1039
[INSERT], Operations, 1000
[INSERT], AverageLatency(us), 78.208
[INSERT], MinLatency(us), 11
[INSERT], MaxLatency(us), 52767
[INSERT], 95thPercentileLatency(us), 62
[INSERT], 99thPercentileLatency(us), 162
[INSERT], Return=OK, 1000
root@shin96:/home/shin96/RocksDB_Festival/YCSB# □
```

Load workload

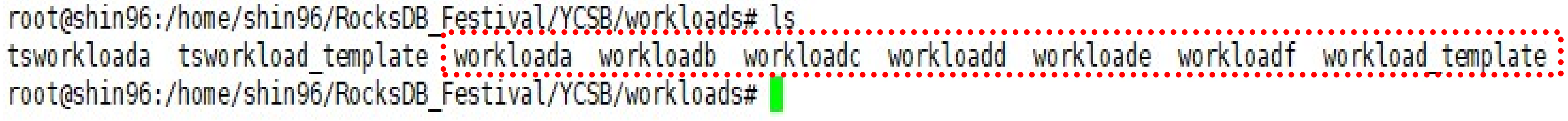


```
Command line: -db site.ycsb.db.rocksdb.RocksDBClient -s -P workloads/workloada -p rocksdb.dir=/tmp/rocksdb-data-ycsb -t
YCSB Client 0.18.0-SNAPSHOT

Loading workload...
Starting test.
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
2021-07-02 22:00:56:630 0 sec: 0 operations; est completion in 0 second
DBWrapper: report latency for each error is false and specific error codes to track for latency are: []
2021-07-02 22:00:56:845 0 sec: 1000 operations; 4366.81 current ops/sec; [READ: Count=521, Max=184, Min=3, Avg=13.95, 90
=28, 99=74, 99.9=140, 99.99=184] [CLEANUP: Count=1, Max=1123, Min=1123, Avg=1123, 90=1123, 99=1123, 99.9=1123, 99.99=112
3] [UPDATE: Count=479, Max=4065, Min=19, Avg=55.03, 90=86, 99=233, 99.9=4065, 99.99=4065]
[OVERALL], RunTime(ms), 229
[OVERALL], Throughput(ops/sec), 4366.812227074236
[TOTAL_GCS_PS_Scavenge], Count, 0
[TOTAL_GC_TIME_PS_Scavenge], Time(ms), 0
[TOTAL_GC_TIME_%_PS_Scavenge], Time(%), 0.0
[TOTAL_GCS_PS_MarkSweep], Count, 0
[TOTAL_GC_TIME_PS_MarkSweep], Time(ms), 0
[TOTAL_GC_TIME_%_PS_MarkSweep], Time(%), 0.0
[TOTAL_GCs], Count, 0
[TOTAL_GC_TIME], Time(ms), 0
[TOTAL_GC_TIME_%], Time(%), 0.0
[READ], Operations, 521
[READ], AverageLatency(us), 13.953934740882918
[READ], MinLatency(us), 3
[READ], MaxLatency(us), 184
[READ], 95thPercentileLatency(us), 41
[READ], 99thPercentileLatency(us), 74
[READ], Return=OK, 521
[CLEANUP], Operations, 1
[CLEANUP], AverageLatency(us), 1123.0
[CLEANUP], MinLatency(us), 1123
[CLEANUP], MaxLatency(us), 1123
[CLEANUP], 95thPercentileLatency(us), 1123
[CLEANUP], 99thPercentileLatency(us), 1123
[UPDATE], Operations, 479
[UPDATE], AverageLatency(us), 55.02713987473904
[UPDATE], MinLatency(us), 19
[UPDATE], MaxLatency(us), 4065
[UPDATE], 95thPercentileLatency(us), 153
[UPDATE], 99thPercentileLatency(us), 233
[UPDATE], Return=OK, 479
root@shin96:/home/shin96/RocksDB_Festival/YCSB# □
```

Run workload

RocksDB Festival

- YCSB
  - ✓ cd workloads
  - ✓ You can control workload files



```
root@shin96:/home/shin96/RocksDB_Festival/YCSB/workloads# ls
tsworkloada  tsworkload_template  workloada  workloadb  workloadc  workloadd  workloade  workloadf  workload_template
root@shin96:/home/shin96/RocksDB_Festival/YCSB/workloads# 
```

Discussion