Week#6 RocksDB Introduction

Sangeun Chae
2018314760

1. INTRODUCTION

RocksDB 는 오픈 소스로 공개한 고속의 쓰기와 읽기에 최적화된 Key-Value 저장소이다. 멀티코어 환경의 SSD 저장 장치 기반의 서버 환경에서 상당한 성능을 보장하는 것이 특징이다. RocksDB 는 Log-structure merge-tree 를 바탕으로 설계된 데이터베이스 엔진이며, LSM tree 구조는, write 를 할 때 append only 방식으로 저장을 하기 때문에, write 를 sequential 하게 처리하여 B+Tree 에 비해 성능이 향상된다. 이번 랩에서는, RocksDB 의 설치와, DB_bench 의 실행을 목적으로 수행될 예정이다.

2. METHODS

Github repository 에서 RocksDB 오픈 소스 코드를 클론 한 후에, 본인의 실험환경에서 RocksDB 를 make 한 후에, DB_bench 를 실행시킨다. DB_bench 를 실행시킨 후, 결과본을 저장한다.

3. Performance Evaluation

3.1 Experimental Setup

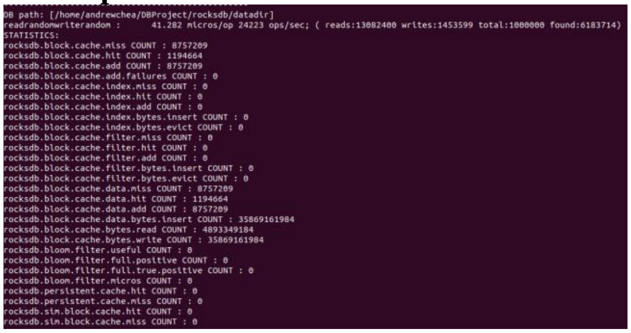
Type	Specification
OS	Ubuntu 20.04.3 LTS
CPU	AMD Ryzen 7 5800X 8-Core Processor (VMware support 4 Core)
Memory	4GB
Kernel	Linux ubuntu 5.11.0.34-generic
Disk	VMware Virtual 80GB

Table 1: System setup

Type	Configuration
Bench Type	“readrandomwriterandom”
Direct flush_compaction	True
Direct read	True
Duration	600

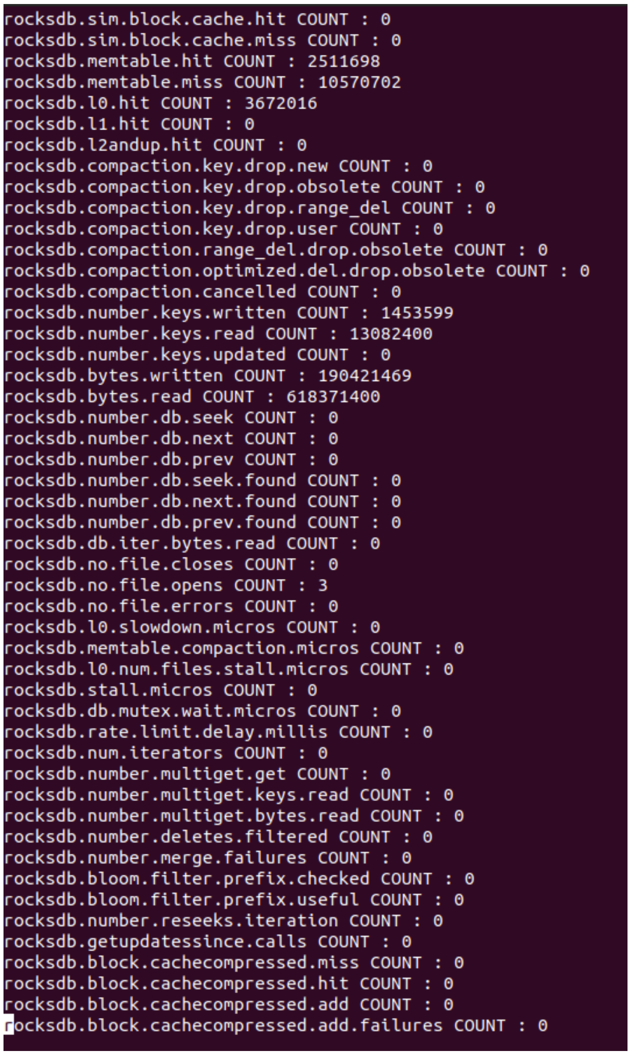
Table 2: Benchmark setup

3.2 Experimental Results



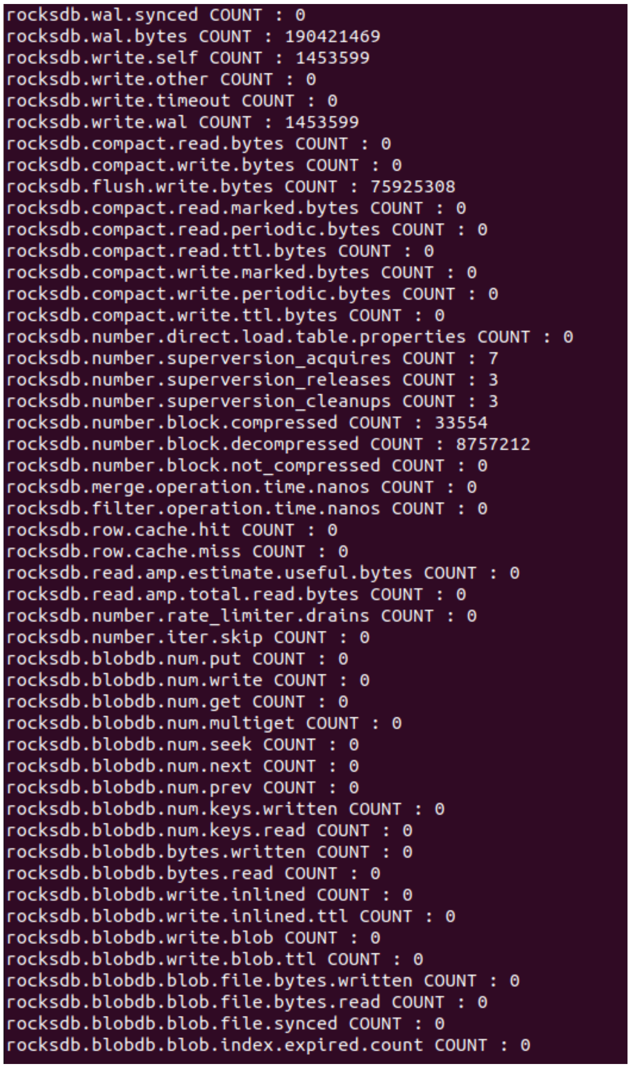
```
DB path: [/home/andrewchea/DBProject/rocksdb/datadir]
readrandomwriterandom :      41.282 micros/op 24223 ops/sec; ( reads:13082400 writes:1453599 total:1000000 found:6183714)
STATISTICS:
rocksdb.block.cache.miss COUNT : 8757209
rocksdb.block.cache.hit COUNT : 1194664
rocksdb.block.cache.add COUNT : 8757209
rocksdb.block.cache.add.failures COUNT : 0
rocksdb.block.cache.index.miss COUNT : 0
rocksdb.block.cache.index.hit COUNT : 0
rocksdb.block.cache.index.add COUNT : 0
rocksdb.block.cache.index.bytes.insert COUNT : 0
rocksdb.block.cache.index.bytes.evict COUNT : 0
rocksdb.block.cache.filter.miss COUNT : 0
rocksdb.block.cache.filter.hit COUNT : 0
rocksdb.block.cache.filter.add COUNT : 0
rocksdb.block.cache.filter.bytes.insert COUNT : 0
rocksdb.block.cache.filter.bytes.evict COUNT : 0
rocksdb.block.cache.data.miss COUNT : 8757209
rocksdb.block.cache.data.hit COUNT : 1194664
rocksdb.block.cache.data.add COUNT : 8757209
rocksdb.block.cache.data.bytes.insert COUNT : 35869161984
rocksdb.block.cache.bytes.read COUNT : 4893349184
rocksdb.block.cache.bytes.write COUNT : 35869161984
rocksdb.bloom.filter.useful COUNT : 0
rocksdb.bloom.filter.full.positive COUNT : 0
rocksdb.bloom.filter.full.true.positive COUNT : 0
rocksdb.bloom.filter.micros COUNT : 0
rocksdb.persistent.cache.hit COUNT : 0
rocksdb.persistent.cache.miss COUNT : 0
rocksdb.sim.block.cache.hit COUNT : 0
rocksdb.sim.block.cache.miss COUNT : 0
```

Figure 1: Experimental result [1]



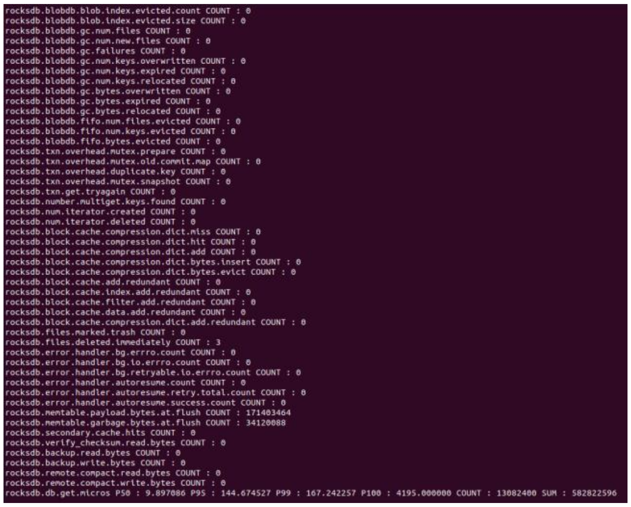
```
rocksdb.sim.block.cache.hit COUNT : 0
rocksdb.sim.block.cache.miss COUNT : 0
rocksdb.memtable.hit COUNT : 2511698
rocksdb.memtable.miss COUNT : 10570702
rocksdb.l0.hit COUNT : 3672016
rocksdb.l1.hit COUNT : 0
rocksdb.l2andup.hit COUNT : 0
rocksdb.compaction.key.drop.new COUNT : 0
rocksdb.compaction.key.drop.obsolete COUNT : 0
rocksdb.compaction.key.drop.range_del COUNT : 0
rocksdb.compaction.key.drop.user COUNT : 0
rocksdb.compaction.range_del.drop.obsolete COUNT : 0
rocksdb.compaction.optimized.del.drop.obsolete COUNT : 0
rocksdb.compaction.cancelled COUNT : 0
rocksdb.number.keys.written COUNT : 1453599
rocksdb.number.keys.read COUNT : 13082400
rocksdb.number.keys.updated COUNT : 0
rocksdb.bytes.written COUNT : 190421469
rocksdb.bytes.read COUNT : 618371400
rocksdb.number.db.seek COUNT : 0
rocksdb.number.db.next COUNT : 0
rocksdb.number.db.prev COUNT : 0
rocksdb.number.db.seek.found COUNT : 0
rocksdb.number.db.next.found COUNT : 0
rocksdb.number.db.prev.found COUNT : 0
rocksdb.db.iter.bytes.read COUNT : 0
rocksdb.no.file.closes COUNT : 0
rocksdb.no.file.opens COUNT : 3
rocksdb.no.file.errors COUNT : 0
rocksdb.l0.slowdown.micros COUNT : 0
rocksdb.memtable.compaction.micros COUNT : 0
rocksdb.l0.num.files.stall.micros COUNT : 0
rocksdb.stall.micros COUNT : 0
rocksdb.db.mutex.wait.micros COUNT : 0
rocksdb.rate.limit.delay.millis COUNT : 0
rocksdb.num.iterators COUNT : 0
rocksdb.number.multiget.get COUNT : 0
rocksdb.number.multiget.keys.read COUNT : 0
rocksdb.number.multiget.bytes.read COUNT : 0
rocksdb.number.deletes.filtered COUNT : 0
rocksdb.number.merge.failures COUNT : 0
rocksdb.bloom.filter.prefix.checked COUNT : 0
rocksdb.bloom.filter.prefix.useful COUNT : 0
rocksdb.number.reseeks.iteration COUNT : 0
rocksdb.getupdatessince.calls COUNT : 0
rocksdb.block.cachecompressed.miss COUNT : 0
rocksdb.block.cachecompressed.hit COUNT : 0
rocksdb.block.cachecompressed.add COUNT : 0
rocksdb.block.cachecompressed.add.failures COUNT : 0
```

Figure 2: Experimental result [2]



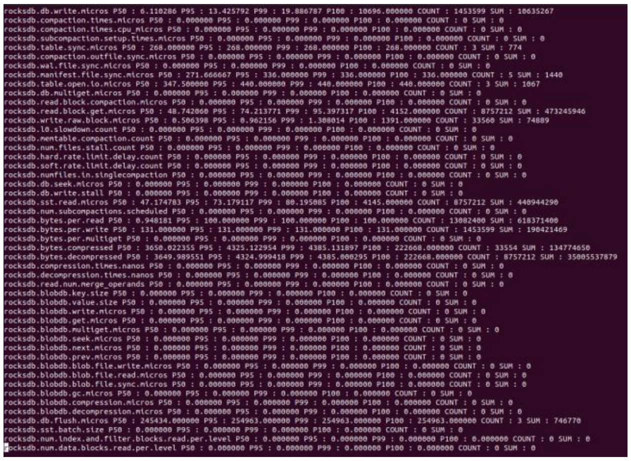
```
rocksdb.wal.synced COUNT : 0
rocksdb.wal.bytes COUNT : 190421469
rocksdb.write.self COUNT : 1453599
rocksdb.write.other COUNT : 0
rocksdb.write.timeout COUNT : 0
rocksdb.write.wal COUNT : 1453599
rocksdb.compact.read.bytes COUNT : 0
rocksdb.compact.write.bytes COUNT : 0
rocksdb.flush.write.bytes COUNT : 75925308
rocksdb.compact.read.marked.bytes COUNT : 0
rocksdb.compact.read.periodic.bytes COUNT : 0
rocksdb.compact.read.ttl.bytes COUNT : 0
rocksdb.compact.write.marked.bytes COUNT : 0
rocksdb.compact.write.periodic.bytes COUNT : 0
rocksdb.compact.write.ttl.bytes COUNT : 0
rocksdb.number.direct.load.table.properties COUNT : 0
rocksdb.number.superversion_acquires COUNT : 7
rocksdb.number.superversion_releases COUNT : 3
rocksdb.number.superversion_cleanups COUNT : 3
rocksdb.number.block.compressed COUNT : 33554
rocksdb.number.block.decompressed COUNT : 8757212
rocksdb.number.block.not_compressed COUNT : 0
rocksdb.merge.operation.time.nanos COUNT : 0
rocksdb.filter.operation.time.nanos COUNT : 0
rocksdb.row.cache.hit COUNT : 0
rocksdb.row.cache.miss COUNT : 0
rocksdb.read.amp.estimate.useful.bytes COUNT : 0
rocksdb.read.amp.total.read.bytes COUNT : 0
rocksdb.number.rate_limiter.drains COUNT : 0
rocksdb.number.iter.skip COUNT : 0
rocksdb.blobdb.num.put COUNT : 0
rocksdb.blobdb.num.write COUNT : 0
rocksdb.blobdb.num.get COUNT : 0
rocksdb.blobdb.num.multiget COUNT : 0
rocksdb.blobdb.num.seek COUNT : 0
rocksdb.blobdb.num.next COUNT : 0
rocksdb.blobdb.num.prev COUNT : 0
rocksdb.blobdb.num.keys.written COUNT : 0
rocksdb.blobdb.num.keys.read COUNT : 0
rocksdb.blobdb.bytes.written COUNT : 0
rocksdb.blobdb.bytes.read COUNT : 0
rocksdb.blobdb.write.inlined COUNT : 0
rocksdb.blobdb.write.inlined.ttl COUNT : 0
rocksdb.blobdb.write.blob COUNT : 0
rocksdb.blobdb.write.blob.ttl COUNT : 0
rocksdb.blobdb.blob.file.bytes.written COUNT : 0
rocksdb.blobdb.blob.file.bytes.read COUNT : 0
rocksdb.blobdb.blob.file.synced COUNT : 0
rocksdb.blobdb.blob.index.expired.count COUNT : 0
```

Figure 3: Experimental result [3]



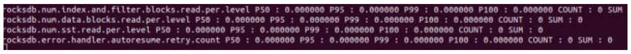
```
rocksdb.blobdb.blob.index.evicted.count COUNT : 0
rocksdb.blobdb.blob.index.evicted.size COUNT : 0
rocksdb.blobdb.gc.num.files COUNT : 0
rocksdb.blobdb.gc.num.new.files COUNT : 0
rocksdb.blobdb.gc.failures COUNT : 0
rocksdb.blobdb.gc.num.keys.overwritten COUNT : 0
rocksdb.blobdb.gc.num.keys.expired COUNT : 0
rocksdb.blobdb.gc.num.keys.relocated COUNT : 0
rocksdb.blobdb.gc.bytes.overwritten COUNT : 0
rocksdb.blobdb.gc.bytes.expired COUNT : 0
rocksdb.blobdb.gc.bytes.relocated COUNT : 0
rocksdb.blobdb.fifo.num.files.evicted COUNT : 0
rocksdb.blobdb.fifo.num.keys.evicted COUNT : 0
rocksdb.blobdb.fifo.bytes.evicted COUNT : 0
rocksdb.txn.overhead.mutex.prepare COUNT : 0
rocksdb.txn.overhead.mutex.old.commit.map COUNT : 0
rocksdb.txn.overhead.duplicate.key COUNT : 0
rocksdb.txn.overhead.mutex.snapshot COUNT : 0
rocksdb.txn.get.tryagain COUNT : 0
rocksdb.number.multiget.keys.found COUNT : 0
rocksdb.num.iterator.created COUNT : 0
rocksdb.num.iterator.deleted COUNT : 0
rocksdb.block.cache.compression.dict.miss COUNT : 0
rocksdb.block.cache.compression.dict.hit COUNT : 0
rocksdb.block.cache.compression.dict.add COUNT : 0
rocksdb.block.cache.compression.dict.bytes.insert COUNT : 0
rocksdb.block.cache.compression.dict.bytes.evict COUNT : 0
rocksdb.block.cache.add.redundant COUNT : 0
rocksdb.block.cache.index.add.redundant COUNT : 0
rocksdb.block.cache.filter.add.redundant COUNT : 0
rocksdb.block.cache.data.add.redundant COUNT : 0
rocksdb.block.cache.compression.dict.add.redundant COUNT : 0
rocksdb.files.marked.trash COUNT : 0
rocksdb.files.deleted.immediately COUNT : 3
rocksdb.error.handler.bg.errro.count COUNT : 0
rocksdb.error.handler.bg.io.errro.count COUNT : 0
rocksdb.error.handler.bg.retryable.io.errro.count COUNT : 0
rocksdb.error.handler.autoresume.count COUNT : 0
rocksdb.error.handler.autoresume.retry.total.count COUNT : 0
rocksdb.error.handler.autoresume.success.count COUNT : 0
rocksdb.memtable.payload.bytes.at.flush COUNT : 171403464
rocksdb.memtable.garbage.bytes.at.flush COUNT : 34120088
rocksdb.secondary.cache.hits COUNT : 0
rocksdb.verify_checksum.read.bytes COUNT : 0
rocksdb.backup.read.bytes COUNT : 0
rocksdb.backup.write.bytes COUNT : 0
rocksdb.remote.compact.read.bytes COUNT : 0
rocksdb.remote.compact.write.bytes COUNT : 0
rocksdb.db.get.micros P50 : 9.897086 P95 : 144.674527 P99 : 167.242257 P100 : 4195.000000 COUNT : 13082400 SUM : 582822596
```

Figure 4: Experimental result [4]



```
rocksdb.db.write.micros P50 : 6.110286 P95 : 13.425792 P99 : 19.886787 P100 : 10696.000000 COUNT : 1453599 SUM : 10635267
rocksdb.compaction.times.micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.compaction.times.cpu_micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.subcompaction.setup.times.micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.table.sync.micros P50 : 268.000000 P95 : 268.000000 P99 : 268.000000 P100 : 268.000000 COUNT : 3 SUM : 774
rocksdb.compaction.outfile.sync.micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.wal.file.sync.micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.manifest.file.sync.micros P50 : 271.666667 P95 : 336.000000 P99 : 336.000000 P100 : 336.000000 COUNT : 5 SUM : 1440
rocksdb.table.open.io.micros P50 : 347.500000 P95 : 440.000000 P99 : 440.000000 P100 : 440.000000 COUNT : 3 SUM : 1067
rocksdb.db.multiget.micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.read.block.compaction.micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.read.block.get.micros P50 : 48.742060 P95 : 74.213771 P99 : 95.397317 P100 : 4152.000000 COUNT : 8757212 SUM : 473245946
rocksdb.write.raw.block.micros P50 : 0.506398 P95 : 0.962156 P99 : 1.308014 P100 : 1391.000000 COUNT : 33560 SUM : 74889
rocksdb.l0.slowdown.count P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.memtable.compaction.count P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.num.files.stall.count P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.hard.rate.limit.delay.count P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.soft.rate.limit.delay.count P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.numfiles.in.singlecompaction P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.db.seek.micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.db.write.stall P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.sst.read.micros P50 : 47.174783 P95 : 73.179117 P99 : 80.195085 P100 : 4145.000000 COUNT : 8757212 SUM : 440944290
rocksdb.num.subcompactions.scheduled P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.bytes.per.read P50 : 0.948181 P95 : 100.000000 P99 : 100.000000 P100 : 100.000000 COUNT : 13082400 SUM : 618371400
rocksdb.bytes.per.write P50 : 131.000000 P95 : 131.000000 P99 : 131.000000 P100 : 131.000000 COUNT : 1453599 SUM : 190421469
rocksdb.bytes.per.multiget P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.bytes.compressed P50 : 3650.022355 P95 : 4325.122954 P99 : 4385.131897 P100 : 222668.000000 COUNT : 33554 SUM : 134774650
rocksdb.bytes.decompressed P50 : 3649.989551 P95 : 4324.999418 P99 : 4385.000295 P100 : 222668.000000 COUNT : 8757212 SUM : 35005537879
rocksdb.compression.times.nanos P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.decompression.times.nanos P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.read.num.merge_operands P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.blobdb.key.size P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.blobdb.value.size P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.blobdb.write.micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.blobdb.get.micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.blobdb.multiget.micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.blobdb.seek.micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.blobdb.next.micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.blobdb.prev.micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.blobdb.blob.file.write.micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.blobdb.blob.file.read.micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.blobdb.blob.file.sync.micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.blobdb.gc.micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.blobdb.compression.micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.blobdb.decompression.micros P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.db.flush.micros P50 : 245434.000000 P95 : 254963.000000 P99 : 254963.000000 P100 : 254963.000000 COUNT : 3 SUM : 746770
rocksdb.sst.batch.size P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.num.index.and.filter.blocks.read.per.level P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.num.data.blocks.read.per.level P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
```

Figure 5: Experimental result [5]



```
rocksdb.num.index.and.filter.blocks.read.per.level P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM
rocksdb.num.data.blocks.read.per.level P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.num.sst.read.per.level P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
rocksdb.error.handler.autoresume.retry.count P50 : 0.000000 P95 : 0.000000 P99 : 0.000000 P100 : 0.000000 COUNT : 0 SUM : 0
```

Figure 6: Experimental result [6]

4. Conclusion

이번 랩에서는, RocksDB 의 설치와 DB_bench 의 실행을 목적으로 두고 진행했기 때문에, 조건의 변화없이 진행했다. 따라서, benchmark 를 실행했을 때, memtable hit count 와 cache hit count, 그리고 throughput 등의 결과 이외에는 유의미한 분석을 할 수 없었다.

5. REFERENCES

[1] https://github.com/meeeejin/SWE3033-F2021/tree/main/week-6