5A-05

# 非同期 IO ライブラリを利用した SSD 向け NoSQL DB「AIODB」の開発と性能評価

圓戸　辰郎†　Hazem Sufian Sulaiman Al-Edaily‡ Leen Raed Sami Hashem‡　檜田　和浩†

東芝メモリ株式会社†　　Princess Sumaya University for Technology‡

## 1. はじめに

近年，情報技術の発展に伴い，様々な情報（データ）が集められ，記憶装置（ストレージ）に蓄積される．そのデータ量は増大を続ける．データは蓄積されるだけではなく，様々な用途に応じて利用される．その際，必要なデータを蓄積・取得するため，様々なデータベースが利用される．特にトランザクションの必要性がない場合には，NoSQL DB が用いられることが多い．このとき，用途によっては，処理速度向上を目的として，In-Memory の NoSQL DB が利用される．ただし，In-Memory の NoSQL DB は蓄積されたデータをすべてメインメモリ上に読み込むことから，HDD 等のストレージを利用する場合に比べ，多くのメインメモリを必要とする．そのため，処理速度が向上することの代償として，HDD 等のストレージよりも高額な費用が発生する．そこで，高額な費用を抑えるため，高速なストレージの利用を考える．近年，ストレージに関する技術の進歩により，SSD を利用することで従来の HDD に比べ，高速なデータの読み書きができるようになってきている．今回，SSD の利用を想定した NoSQL DB を開発することで，費用対効果の高いストレージシステムを提案する．

本稿では，SSD の利用を想定した NoSQL DB の開発，および，開発した NoSQL DB の性能評価の結果について述べる．なお，開発した NoSQL DB を AIODB と呼ぶ．

## 2. In-Memory NoSQL DB

本稿では，サーバ 1 台における SSD の利用を想定している．つまり，ローカルストレージにデータを蓄積する DB を前提としている．その前提で，本節では OSS の In-Memory の NoSQL DB を評価し，ボトルネックを調査する．本稿において詳細は割愛するが，いくつかの OSS の In-Memory の NoSQL DB(LMDB[1], MDBM[2], Redis[3]等)について調査を行った．本稿では特に LMDB について詳細を述べる．LMDB を選択した理由は，前述の中で最も良い性能を示したからである．LMDB の性能評価として，表 1 の条件を基に評価用プログラムを作成した．このプログラムは，次の手順を key-value のペア数だけ繰り返すものである．

1) 与えられた key から value を検索
2) 得られたポインタからデータをコピー

本評価では，手順 2)のコピーするサイズを変更し，処理時間にどのような影響が現れるか確認した．評価に必要な key-value は，プログラム評価前に事前に用意し，メインメモリに読み込んでおく．なお，value は適当な文字列である．評価に用いた環境を表 2 に示す．また，評価の結果を図 1 に示す。コピーサイズが 0 [byte]の場合は，コピーを行っていないことを示す．その際には 1M [IOPS]の性能になっているが，コピーサイズが 4K [byte]の場合には，256K [IOPS]まで下がっている．このことから，key に対応する value の検索処理よりも，value のコピーが性能上のボトルネックになっていることがわかる．

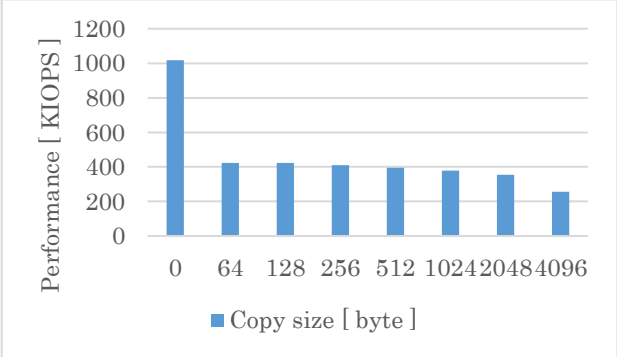


図 1 Evaluation of LMDB

表 1 Conditions of LMDB's index

| | |
|---|---|
| Key size | 8 [byte] |
| Value size | 4096 [byte] |
| The number of key-value pairs | 1 million |

表 2 Specifications of the Server

| | |
|---|---|
| CPU | Intel i5-6500 : 3.2 [GHz] |
| Memory | DDR4 2133 [MHz] : 64 [GB] |

Development and Evaluation of NoSQL DB "AIODB" for SSDs with Asynchronous IO Library

†Tatsuro Endo, Kazuhiro Hiwada: Toshiba Memory Corporation
‡Hazem Sufian Sulaiman Al-Edaily, Leen Raed Sami Hashem: Princess Sumaya University for Technology

## 3. SSD向けNoSQL DB

### 3.1 システム構成

本小節では，今回開発したSSD向けNoSQL DBの構成について述べる．ここでは，実データはSSDに保存し，そのデータ位置を既存のNoSQL DBで管理する方法を提案する．また，データ取得時の負荷軽減のため，非同期IOライブラリを利用する．開発したNoSQL DBは，次のように2つのライブラリからなる．

1) LMDB: データ位置を管理
2) libaio[4]: データをSSDから非同期に取得

1)によって，あるkeyが与えられれば，SSD上のどの位置にvalueが保存されているか検索できる．2)によって，得られたvalueの位置からデータを非同期処理によって取得できる．

### 3.2 API

本小節では，開発したNoSQL DBのAPIについて述べる．次の通り，NoSQL DBとして基本的なAPIを実装している．

1) open: 処理開始
2) close: 処理終了
3) get: keyからvalueを取得
4) put: key-valueの保存

利用方法として，valueを取得したい場合は，1)->3)->2)の流れで行う．また，key-valueを保存したい場合は，1)->4)->2)の流れで行う．

## 4. 性能評価

本節では，開発したNoSQL DBの評価について述べる．評価プログラム，および条件は前述の2節と同様である．使用したSSDはIntel Optane SSD 900P[5]である．この評価ではデータのコピーサイズは4K [byte]の固定値である．これは，SSDが扱うブロックサイズが4K [byte]であり，それより小さいサイズを指定できないからである．評価結果を図 2 に示す．この評価では，libaioのキュー数を変更し，その様子を見た．キュー数が64のあたりで飽和しており，およそ370K [IOPS]となっている．それ以降，緩やかに数値が増え，キュー数が1024のときに，およそ375K [IOPS]となった．キュー数はある程度大きい数値としたほうがよいが，大きくし過ぎても性能はあまり変化しない．なお，fio[6]による同等の条件を用いた測定では377K [IOPS]であり，AIODBのライブラリとしてのオーバーヘッドは少ないと言える．また，これまでの評価を比較すると，図 1 のvalueのコピーサイズが1K [byte]の場合と，図 2 の飽和時の性能がほぼ同等となっている．つまり，1K [byte]のデータをDBから取得する場合，In-MemoryのNoSQL DBであるLMDBとSSDから直接データを転送するAIODBが同等の性能になっており，SSDに対するDRAMの性能の優位性が失われていると言える．1K [byte]より大きいサイズのデータを扱う場合には，In-MemoryのNoSQL DBより我々のAIODBのほうがよい性能となっている．非同期IOライブラリによるCPUの負荷軽減，およびSSDから直接データ転送を行うことによるCPUの負荷軽減の複数の効果によって，このような結果が得られたと思われる．

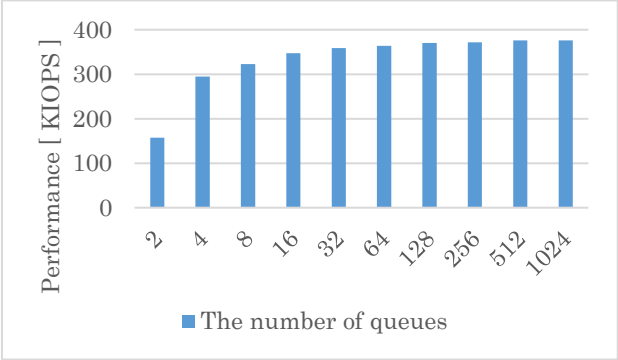


図 2 Evaluation of NoSQL DB:AIODB

## 5. おわりに

本稿では，既存のIn-MemoryのNoSQL DBの性能上のボトルネックを調査し，その改善のため，非同期IOライブラリを用いて，SSD向けNoSQL DB「AIODB」を開発した．前述の評価において示したとおり，条件次第ではIn-MemoryのNoSQL DBに迫る，もしくは上回る性能を発揮することを確認した．これにより，費用対効果が高いNoSQL DBを提案することができた．

今後は，さらなる性能向上を目指し，複数台のSSDを利用するなどの対応を行っていきたい．また，ユースケースを考慮した評価を行うことで，様々な条件に対応させていきたい．さらに，SSD向けに開発された他のNoSQL DBとも比較したい．

## 参考文献

[1]LMDB: https://github.com/LMDB/lmdb
[2]MDBM: https://github.com/yahoo/mdbm
[3]Redis: https://redis.io/
[4]libaio: http://lse.sourceforge.net/io/aio.html
[5]Intel Optane SSD 900P: https://www.intel.co.jp/content/www/jp/ja/solid-state-drives/optane-ssd-900p-brief.html
[6]fio: https://github.com/axboe/fio