

# Efficiency of KV read and write on different KV implementations

---

Author: Zhoujie(Jason) Ding

## Abstract

---

There are many kinds of KV database to use for our kaas, and we are looking for the one that performs best. In my benchmark, I simply set the shortest time for KV read and write as the metric. I currently want to compare the performance between Redis and Anna. In this report, I have three KV databases: Redis, Anna, and the local database acting like a dictionary. My inputs are different sizes of matrices, varying from 5 by 5 to 5000 by 5000. In addition, different tests are performed using these matrices for different types of KV read and write in order to simulate the real world situations.

## Methods

---

I use the `time` function in python api to measure the duration of reading or writing the matrices in the KV database. I also use the `testenv` function in `utils.py` to invoke and close the Redis or Anna so that I reset the server in every test. Moreover, I repeat every test multiple times and record every time slot in a list. Then, I calculate the mean and the standard deviation of the time slots in order to reduce error.

Below is the explanation for each function I use in `kvBench.py`.

`test_putAndget(kv, k, v, is_R, is_W)`: Function for KV read and write. It is the function that I test for time.

`setup_mode(mode)`: Function to set up the mode of KV database.

`test_mixed(mode, klst, vlst, is_cold, is_R, is_W)`: Test time for mixed sizes of matrices.

`test_not_mixed(mode, klst, vlst, is_manyRW, is_cold, is_R, is_W)`: Test time for different sizes of matrices one size at a time.

`time_putAndget(mode, is_mixed=False, is_manyRW=False, is_cold=True, is_R=True, is_W=True)`: Main function I use to test time with different hyper parameters.

Also, it's important to explain the parameters.

`is_mixed`: whether I test with different sizes of matrices in a single read or write.

`is_manyRW`: whether I test with multiple reads and writes at the same time with one size of matrices.

`is_cold`: whether I test without warming up the cache, meaning whether I fill in the cache with some data of matrices before beginning my test.

`is_R`: whether I test reads.

`is_W`: whether I test writes.

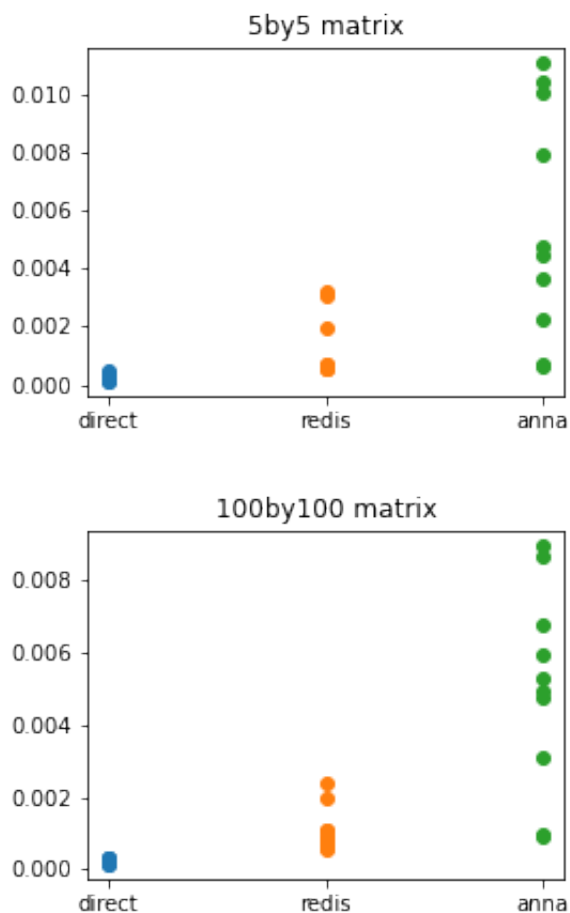
## Results

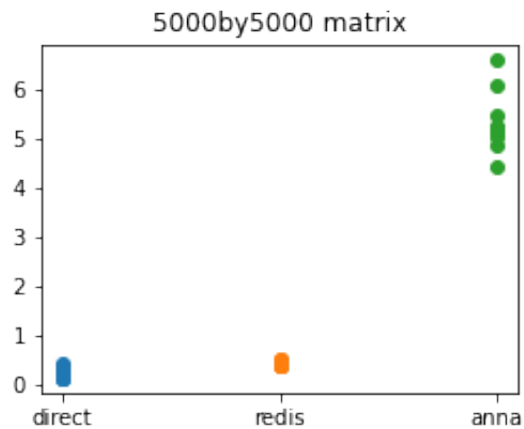
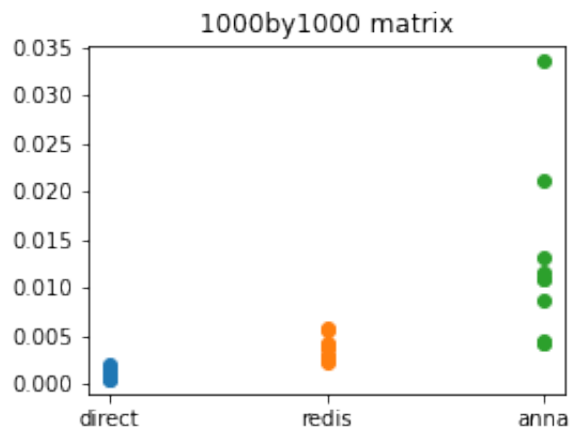
---

The default setting for all these tests is: `is_mixed=False`, `is_manyRW=False`, `is_cold=True`, `is_R=True`, and `is_W=True`.

### 1. Test cold and warm caches

I only change `is_cold`. Below is the plot comparing cold and warm caches before testing on reads and writes.





## 2. Test mutiple reads and writes

# Conclusion

---

## 1. Test cold and warm caches

According to the plots, I can see that as the size of the matrix increases, Redis behaves much better than Anna, and Redis acts nearly as well as the local dictionary. One thing to notice is that for 5000 by 5000 matrix, Anna performs much worse.

## 2. Test mutiple reads and writes