Reproducibility Report for ACM SIGMOD 2020 Paper: “Timely Reporting of Heavy Hitters using External Memory”

EDDY GHABACH, EURECOM
RAJA APPUSWAMY, EURECOM

The paper presents external memory algorithms for the Timely Event Detection problem. The paper shows that these algorithms perform much less than one I/O per query and are limited only by I/O bandwidth. The reproducibility fulfills this objective, since it produces results similar to the paper results.

1 INTRODUCTION

The reproduced paper [1] is written by Pandey and Singh *et al.* from Carnegie Mellon University and Wellesley College respectively. The paper shows how to adapt heavy-hitters algorithms to external memory to solve the Timely Event Detection problem on large high-speed streams. The reproducibility shows similar results to the ones displayed on the paper.

2 SUBMISSION

The source code, latex code for building the paper and the scripts are available on: https://zenodo.org/record/4380373#.X9_KcHVKiV4. The downloaded file contains two directories:

- **LERT-src**: contains the source codes and the scripts. A directory called “sigmod20_raw", initially empty, will contain the generated data after running the scripts.
- **SIGMOD20-paper**: contains tex files to generate the pdf of the paper. The data generated in the “sigmod20_raw" directory is used to create the graphs of the paper.

The generated pdf of the paper after running the scripts is available on: https://gitlab.eurecom.fr/ghabach/sigmod

3 HARDWARE AND SOFTWARE ENVIRONMENT

The paper experiments were done on two machines. Timeliness and I/O performance benchmarks were run on a machine, and scalability benchmarks were run on another machine. The reproducibility experiments were run on a single machine. Table 1 shows hardware and software specifications of those machines.

Table 1. Hardware & Software environment

	Paper [a]	Paper [b]	Repro Review
CPU	Intel i7-6700HQ	Intel E5-2683 v4	Intel i9-10920X
cores	4	64	12
GHz	2.60	2.10	3.50
RAM	32GB	512GB	132GB
Storage	1TB SSD	1TB SSD	1TB SSD
OS	Ubuntu 18.04	Ubuntu 18.04	Ubuntu 18.04

[a] timeliness and I/O performance benchmarks
[b] Scalability benchmarks

4 REPRODUCIBILITY EVALUATION

4.1 Process

The process for reproducibility started by running the “limitMem.sh" script to setup cgroups profile. Next the “run_experiements.sh" script is run to generate the data inside the “sigmod20_raw" directory. This process takes several hours. The inital author of the paper was contacted to solve some errors that were produced when running the scripts, and this is due to some permission issues on the machine on which the reproducibility experiments were taking place. Finally, to generate the pdf of the paper, the “make" command was run inside the “SIGMOD20-paper" directory.

4.2 Results

The reproducibility shows that all findings of the paper were reproduced and showing similar results to the ones displayed on the paper. The distribution of count stretch and time stretch of different data structures (as shown in Figure 1), is similar in both the paper and the reproducibility result. The distribution of time stretch for different distributions (Figure 2.a) shows similar results, with a small variation of the Round-robin values. The distribution of count stretch with different buffering strategies shows similar results (as shown in Figure 2.b). The distribution of count stretch vs lifetime (Figure 3) allows similar interpretation when comparing the results obtain in the paper and the reproducibility results. The total read and write I/O performed are similar in the paper and the reproducibility results (as shown in Figure 4). Similarly, the items inserted per second (Figure 5.a), the insertion throughput with increasing number of threads (Figure 5.b), and the instantaneous throughput (Figure 5.c) show identical values in the paper and the reproducibility results.

5 SUMMARY

The evaluation of the reproducibility results shows that it produces similar results compared with the paper results.

REFERENCES

[1] Prashant Pandey, Shikha Singh, Michael A Bender, Jonathan W Berry, Martín Farach-Colton, Rob Johnson, Thomas M Kroeger, and Cynthia A Phillips. 2020. Timely Reporting of Heavy Hitters using External Memory. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*. 1431–1446.