Reproducibility Report for ACM SIGMOD 2021 Paper: "Maximizing Persistent Memory Bandwidth Utilization for OLAP Workloads"

EUGENIO MARINELLI, EURECOM, France
RAJA APPUSWAMY, EURECOM, France

The paper systematically evaluates Intel's Optane DIMMs as the first commercially available PMEM and provides 7 best practices for its use in OLAP workloads. The paper shows that PMEM is a suitable alternative to expensive and capacity-limited DRAM for OLAP workloads. The reproducibility fulfills this objective since it produces results similar to the paper results.

1 INTRODUCTION

The reproduced paper [1] is written by Björn Daase, Lars Jonas Bollmeier, Lawrence Benson, and Tilmann Rabl from the University of Postdam. The paper shows that PMEM is suitable for data-intensive systems and provides a higher capacity at lower prices than DRAM while performing only slightly worse. Overall, the reproducibility shows similar results to the ones displayed on the paper. We are able to reproduce all the experiments reported in the paper except for the left graph in "Figure 14: Star Schema Benchmark Performance".

2 SUBMISSION

The source code and the scripts for reproducing the paper results are available on GitHub[1]. The downloaded file contains 4 main directories:

- **microbenchmarks**: contains the source code for the performed microbenchmarks.
- **plot_script**: contains the scripts that generate the graphics in the paper.
- **script**: contains the script for running all the executables and gather the results.
- **ssb**: contains the source code for Star Schema Benchmark experiments.

3 HARDWARE AND SOFTWARE ENVIRONMENT

Table 1 shows the hardware configuration used for the paper and the reproducibility. As we did have the required hardware for reproducing the paper, we worked with the authors in getting remote access to their hardware testbed. Thus, the configuration used by the authors in the paper matches the one used for reproducibility. All benchmarks are written in C/C++ and assembly.

Table 1. Hardware environment

	Paper
CPU	Intel® Xeon® Gold 5220
cores	18
DRAM	6 GB x 6 channels x 2 sockets
PMEM	128 GB x 6 DIMMs x 2 sockets
SSD	Intel® DC P4610 Serie

[1] https://github.com/hpides/pmem-olap.git

4 REPRODUCIBILITY EVALUATION

4.1 Process

The process for reproducibility started by getting access to the authors' hardware infrastructure. We were able to get a fully configured instance on which the GitHub [2] repository with the required environment (paths, tools, etcetera) set and all the executables readily compiled. First, we ran the `microbenchmark.sh` script in order to gather all the results corresponding to the microbenchmark evaluation in the paper (from Figure 3 to Figure 13 in the paper). Then, we ran the script run_ssb.sh in order to gather the results for Star Schema Benchmark (SSB). We ran into several issues during this process. First, the whole benchmarking process took more than 72 hours. The first issue we ran into was getting node reservations for such a long duration. Second, we found that despite preconfiguration, their cluster nodes lacked certain Python libraries that were required for plotting. Thus, we had to manually download the generated results and plot them locally. We also ran into several errors while running the script that was caused due to permission issues on the machine on which the reproducibility experiments were taking place. Through iterative rounds of communication with the authors, we were able to fix all the problems and solve these errors.

4.2 Results

The reproducibility shows that all findings of the paper were reproduced and the results are similar to the ones displayed on the paper. The trend displayed in the reproduced Figures [3,4,5,6 7,9,10,11,12,13] is very close to the one shown in the paper. Table 1 also shows the same behavior as the one in the paper. Concerning the bar charts in Figure 14 in the reproduced paper, we are able to reproduce only the graph on the right (Handcrafted C++ sf=100). We were unable to reproduce the graph on the left as we ran into some unsolved issues in accessing the required data. Due to time constraints, the authors were not able to fix the problem. However, we believe that this is not a major problem given that the rest of the paper was reproducible to a large extent, thus, validating the key findings from the paper.

5 SUMMARY

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

REFERENCES

[1] Björn Daase, Lars Jonas Bollmeier, Lawrence Benson, and Tilmann Rabl. 2021. Maximizing Persistent Memory Bandwidth Utilization for OLAP Workloads. In *Proceedings of the 2021 International Conference on Management of Data (SIGMOD '21), June 20–25, 2021, Virtual Event, China.* ACM, Virtual Event, China.

[2] https://github.com/hpides/pmem-olap.gi