

Reproducibility Report for ACM SIGMOD 2020 Paper: “BinDex: A Two-Layered Index for Fast and Robust Scans”

JIASHEN CAO, Georgia Institution of Technology, USA

We can reproduce most results described in the BinDex paper and the premise of the paper is reproducible.

1 INTRODUCTION

This reproducibility report is for paper “BinDex: A Two-Layered Index for Fast and Robust Scans” [1]. The paper is affiliated with Fudan University. We can reproduce most results presented in the paper, but performance results about Zipf distribution cannot be reproduced. During reproducing process, authors were responsive to offer help for debugging or improving their code.

2 SUBMISSION

Authors provide a zip file, which contains following items:

- script to run experiments
- script to compile and run baseline experiments
- script to generate necessary data
- script to produce figures based on generated data

We think the provided zip file is sufficient to run all experiments.

3 HARDWARE AND SOFTWARE ENVIRONMENT

Table 1. Hardware & Software environment

	Paper	Repro Review
CPU	Intel Xeon E5-2695	Intel Xeon E5-2696
cores	18	44
GHz	2.1	2.2
RAM	128GB	256GB
L3	45MB	55MB
OS	Ubuntu-18.04	Ubuntu-16.04
g++	6.5	5.4

4 REPRODUCIBILITY EVALUATION

4.1 Process

The entire process takes about two monthes, which include the time that authors help to improve the code and script. The code initially contains bugs, which include path mismatch, wrong script command, in-compatible compilation issues. With the help from authors, the script command is fixed, but there still exists some path mismatch and in-compatible compilation issues as we write this report. Another issue is experiments about BinDex performance on Zipf distribution take much longer than expected. Authors mention it should finish within one day, but the experiment cannot finish even after two days. We have requested authors to fix, but the updated code provided by authors still has the same issue.

4.2 Results

We can now successfully reproduce Figure 5, 6, 7, and 8. As the paper demonstrates, our results also confirm that BinDex outperforms other baseline on uniform distribution. The relative speedup also follows the main premise of the BinDex paper. As we discussed, we cannot reproduce Figure 9 (the performance comparison of BinDex with other baseline on Zipf distribution). Consequently, due to missing data, Figure 10 and 13 cannot be reproduced as well. We can reproduce Figure 11, 12, and 14, which also confirms the results presented in the paper.

REFERENCES

- [1] Linwei Li, Kai Zhang, Jiading Guo, Wen He, Zhenying He, Yinan Jing, Weili Han, and X. Sean Wang. 2020. BinDex: A Two-Layered Index for Fast and Robust Scans. In *SIGMOD*. 909–923.