Readme for reproducibility submission of SIGMOD'21 paper ID 68

1 Source code info

Repository: https://github.com/Yanqing-UTAH/ATTPCode
Programming Langauge: mainly C/C++, some of the scripts are written in Python
Required software/library: gcc/g++ >= 8 (requires support for -std=gnu++17), make, lapack and lapacke, blas and cblas, fftw3, python3 and the common shell programs (bash, grep, sed, and etc.).
Optional software/library: for plotting figures: jupyter notebook, python2 matplotlib, pandas, numpy; for recreating "configure" script if that does not work in your environment: m4, autoconf, autoheader.
Hardware requirement: we assume the architecture implements <= 48-bit virtual address space and always uses x86-64 canonical addresses. That's the case with any Intel/AMD processor other than Intel Ice Lake (which supports 5-level paging) as of right now (2021).

2 Test environment

Software environment:

- Ubuntu 18.04.6 LTS
- GNU make 4.1
- GCC/G++ 8.3.0
- liblapack-dev 3.7.1
- liblapacke-dev 3.7.1
- libblas-dev 3.7.1
- libatlas-base-dev 3.10.3
- libfftw3-dev 3.3.7

Note that these are the ones installed at the time we performed the experiments but it should be ok if you use newer versions. If you're using Ubuntu, you should be able to use the following to get all these packages:

```
sudo apt install gcc g++ make liblapacke-dev libatlas-base-dev libfftw3-dev
```

Hardware environment:

Node type	1	2
CPU	Intel Core i7-3820	Intel Xeon E5-1650 v3
CPU Frequency	3.6 GHz	3.50 GHz
L1 Cache	32KB + 32 KB	32KB + 32 KB
L2 Cache	256 KB	256 KB
L3 Cache	10 MB (shared)	15 MB (shared)
Memory	64GB (DDR3-1600 x 8)	128GB (DDR4-2133 x 4)
Secondary storage	WD HDD 7200RPM 2TB	Seagate HDD 7200RPM 1TB
Network	not used	not used

The experiments are single-threaded and we only launch one experiement at a time on a single node so the number of cores or hyperthreading is irrelevant to our purpose. Cache size is also irrelevant for our purpose as our data structure size far exceeds even the L3 cache size (listed below just for reference). On the other hand, we have 10 type-1 nodes and 6 type-2 nodes and we may launch any of the experiments on any one of them depending on the availability. We made sure that there were no computation or I/O heavy programs running concurrently.

Preparing datasets