**Database Project (SWE3033) (Fall 2023)**

**Homework #2 (100pts, Due date: 09/13)**

**Student ID**: 2020315798

**Student Name**: Choi Jin Woo

1. **[10pts]** Write the system setup of your environment. Fill in the blank below.

|  |  |
| --- | --- |
| **Type** | **Specification** |
| OS | Ubuntu 18.04 |
| CPU | Intel(R) Core(TM) i5-8265U CPU @ 1.60GHz (2 Cores) |
| Memory (RAM) | 6 GB |
| Kernel | Linux version 5.4.0-84-generic |
| Data Device (Optional) |  |
| Log Device (Optional) |  |

2. **[10pts]** Write the benchmark setup of your environment. Fill in the blank below.

|  |  |
| --- | --- |
| **Type** | **Specification** |
| DB Size | 2GB |
| Buffer Pool Size | 1GB |
| Benchmark Tool | tpcc-mysql |
| Runtime | 1200s |
| Connections | 8 |

3. **[50pts]** Assume that you have a database with a size of 10 warehouses. Determine the **optimal percentage of buffer pool size to maximize TPC-C benchmark performance**. Present experimental results and analyze the results yourself. You must include **1. chart** **of TpmC and hit ratio**, **2. result of least three buffer pool sizes**. Explain the reason for the performance shift as the buffer size varies. (The buffer pool size should not exceed the database size.)

|  |
| --- |
| I have conducted 3 different buffer pool sizes: 256MB (25%), 512MB (50%), 768MB (75%)   1. **256 MB**   Buffer Pool Hit Rate: 48.3%  TpmC: 2574.35       1. **512 MB**   Buffer Pool Hit Rate: 57.5%  TpmC: 3734.2       1. **768 MB**   Buffer Pool Hit Rate: 78.8%  TpmC: 4202.7      As a result, a chart goes as follows:  Conclusion:  Increasing the buffer pool size is expected to lead to an increase in TpmC (Transactions per Minute in the TPC-C benchmark) and the Buffer Hit Rate. As the allocated buffer size increases, the capacity to cache data grows, resulting in a higher hit rate. Consequently, this is likely to lead to an increase in the number of transactions processed per minute.  I couldn’t attach the overall graph of the log files, because Gnuplot mentioned in the Github repository referred in the course slides does NOT work somehow. I followed the instructions and modified with 3 input text files from log files, however, it didn’t work with even 2 inputs also. |

4. **[30pts]** Why is it difficult to store the entire database in memory in real world, and why do we need to use buffer pool in database management system?

|  |
| --- |
| Storing an entire database in memory can be challenging for the following reasons.   * Databases tend to grow as more data is added. So it will eventually run out of memory. * Many databases are massive in size, ranging from gigabytes to terabytes or even more. * Storing the entire database in RAM makes it vulnerable to data loss in case of power outages or system crashes. * RAM is more expensive than HDD or SSD. To store an entire database in RAM, it is required to invest significantly in high-capacity RAM modules, increasing the overall cost of the system.   As a result, with the above reasons, it is difficult to store the entire database in memory in real world.  While there are some difficulties, we use buffer pool for the following reasons.   * Most commonly accessed data is available in memory, providing fast access times. * By caching data in memory, buffer pools reduce the need to read data from slower storage devices like hard drives or SSDs. * It is cost-effective compared to the amount of RAM usage storing the entire database, since it uses only a smaller amount of RAM. * Data is persistently stored on on-volatile storage, so ensures the durability. |