COMP4801 – Final Year Project

Benchmarking Hybrid Transactional and Analytical Processing Graph Database Systems

Project Plan

Srivastava Dhruv
Department of Computer Science

Supervised by
Dr. Heming Cui
Department of Computer Science

October 2, 2022
## Contents

1 Introduction ............................................................................................................................ 3

2 Background ........................................................................................................................ 3  
   2.1 Online Transaction Processing (OLTP) ................................................................. 3  
   2.2 Online Analytical Processing (OLAP) ................................................................. 4  
   2.3 Hybrid Transactional and Analytical Processing (HTAP) ................................. 4  
   2.4 Online Social Trading ......................................................................................... 4

3 Objective ............................................................................................................................ 5  
   3.1 Scope ..................................................................................................................... 5  
   3.2 Deliverables .......................................................................................................... 6

4 Methodology ..................................................................................................................... 6

5 Risks, Challenges, and Mitigation ................................................................................... 7

6 Project Milestones .......................................................................................................... 8

7 Project Schedule ............................................................................................................ 8

8 Conclusion ....................................................................................................................... 9

9 References ....................................................................................................................... 11
1 Introduction

Graph database systems are a relatively new type of database systems that have been gaining popularity in recent years. These systems offer significant advantages over traditional relational database systems when supporting modern agile software development as the data schema of the data being stored can be changed flexibly depending on the business requirements [1]. Graph databases generally offer better performance than relational databases when querying growing data as unlike relational databases, they do not perform table joins to compute query results [1].

Database systems support different kinds of workloads and can, therefore, be categorized into two types: Online Transactional Processing (OLTP) database systems and Online Analytical Processing (OLAP) database systems. Some systems support both OLTP and OLAP workloads simultaneously, and these are called Hybrid Transactional and Analytical Processing (HTAP) database systems [2, 3, 4, 5]. HTAP database systems are a good alternative for architectures containing the Extract-Transform-Load (ETL) process, where data from an OLTP database system is transferred to an OLAP database system for data analytics, providing advantages for certain applications [3, 4].

HTAP Graph Database Systems Research is an emerging field in academia and only a few contributions in this field have been made in the recent years. One of the key challenges that contributors in this field face is the absence of any standard benchmarks for HTAP graph database systems, which makes it difficult to compare the performance of different proposed architectures. In order to address this identified research gap, a new benchmarking framework will be designed and implemented in this final year project (FYP). The benchmarking framework will simulate an online social trading application and generate the performance metrics for the system under test (SUT). The online social trading domain was chosen for this benchmarking framework as no benchmark for this problem domain exists, and thus be an additional contribution.

The rest of the project plan is structured in the following way: section 2 provides background on some terms used throughout this document; section 3 contains the objectives for this project; section 4 elaborates on the methodology employed for the development and evaluation of the benchmarking framework throughout this project; section 5 lists some challenges that are expected to be encountered during the course of this project and some plans for their mitigation; section 6 contains the detailed project schedule; and section 7 lists the project milestones and the expected learning hours for each milestone.

2 Background

This section will provide some background on some terms used throughout this project plan.

2.1 Online Transaction Processing (OLTP)
In an OLTP workload, the queries to the database system are uncomplicated and are transactional, requiring ACID (Atomicity, Consistency, Isolation, Durability) support [2, 6]. Systems supporting this workload type are designed to process a lot of transactions simultaneously [2, 6]. The transactional queries are write-intensive, and usually insert new data, update, or delete existing data [2, 6]. This workload type requires high availability, quick transaction processing times and data consistency [2].

2.2 Online Analytical Processing (OLAP)

An OLAP workload consists of read-only queries [2, 6]. The queries in this workload type are complicated as they consist of massive amount of data reads to perform complex analysis on the data being stored in the database system [2, 6]. This workload type is resource-intensive as massive amounts of data is processed in each query and requires quick read times [2].

2.3 Hybrid Transactional and Analytical Processing (HTAP)

Since OLTP and OLAP workloads have competing requirements, one requiring high availability and the other being resource intensive [2], applications requiring both workloads simultaneously were generally handled by having both an OLTP system and an OLAP system to get the best performance from each system [4]. The OLTP system would serve the transactional workload and transfer the data to an OLAP system from time to time for analytics, with the help of the ETL process mentioned previously [4, 5]. The biggest disadvantage of this architecture is that the data available for analytics is usually stale [4], making it unsuitable for applications requiring real-time analytics.

HTAP database systems aim to serve both OLTP and OLAP workloads simultaneously. These systems make fresh data to be available for analytics, allowing real-time analytics on incoming data [4]. They also reduce management costs as there is no need for an ETL process and reduce storage costs as there is no need to maintain two copies of data, one in the OLTP system and the other in the OLAP system [4].

2.4 Online Social Trading

Online social trading is essentially a social media platform where users can trade on the stock market publicly within the platform and perform other interactions to gain more followers [7, 8, 9]. Most platforms allow followers to automatically duplicate the trades of the investor they are following and the investor, in turn, has the chance to earn a commission from their followers’
automatically duplicated trades [7, 8, 9], therefore, offering benefits to both followers and leaders. Some popular online social trading platforms are eToro, and Zulutrade.

This domain is suitable for graph database systems as the data schema contains relationships like investor A ‘follows’ investor B and investor A ‘copies trade made by’ investor B, and most actions, like making a trade, happen around these relationships. A relational database system may offer poorer performance than a graph database system as the data size grows due to the need to compute ‘table joins’ for such operations, as mentioned previously. This domain is also suitable for HTAP systems as it requires OLTP capabilities for actions like making a trade and may benefit from real-time analytics made possible by HTAP systems. Some examples of such analytics can be prediction of stock trends based on the actions being made on the platform and recommendations for investors on who to follow/copy.

3 Objective

This section defines the objective of this project, the scope of work, and expected deliverables. As stated previously, the objective of this project is to develop a benchmarking framework for HTAP graph database systems to address the research gap of absence of such benchmarks. It is hoped that with this benchmark, comparison of different HTAP graph database architectures will become easier for the contributors in this field. The benchmark targets the online social trading problem domain and aims to simulate a workload for the domain. This will address the absence of benchmarks for this problem domain. The benchmark will also be designed as a framework, so that other problem domains and databases supporting different query languages can potentially be benchmarked using the components listed in section 3.1.1.

3.1 Scope

This section describes the provisional scope of this project.

3.1.1 Benchmarking Framework Development

A novel benchmarking framework for HTAP graph database systems will be designed and implemented during the project. The benchmarking framework is expected to target the online social trading domain and will evaluate the performance of the SUT under such workload. The benchmark will contain: a base data generator, a workload generator, transactional and analytical query drivers, and a post-benchmark processor.

The purpose of each benchmark component is discussed below:

- **Base Data Generator:** To generate the base data for the problem domain.
- **Workload Generator:** To generate the query workload for benchmarking.
• **Transactional Query Driver:** To send transactional queries and collect data for evaluating the performance of the SUT.
• **Analytical Query Driver:** To send analytical queries and collect data for evaluating the performance of the SUT.
• **Post-Benchmark Processor:** To evaluate the performance of SUT for different metrics, based on the data collected by the query drivers.

### 3.1.2 Evaluation

The benchmarking framework will be used to evaluate some HTAP graph database systems or related systems available for benchmarking. There are some risks and challenges associated with this part of the project that have been discussed in section 5.1.

### 3.2 Deliverables

The tentative project deliverables are as follows:

• An online social trading benchmarking framework for HTAP graph database systems with the following components:
  o Base Data Generator
  o Workload Generator
  o Transactional Query Driver
  o Analytical Query Driver
  o Post-Benchmark Processor
• Evaluation of HTAP graph database related systems.

### 4 Methodology

This section describes the methodology to be employed for the different phases of the project. The project consists of two phases: Benchmarking Framework Development, and Evaluation.

#### 4.1 Benchmarking Framework Development

In order to design the benchmark to model the online social trading problem domain, a literature review will be conducted to understand the data schema, query types, and query distribution for the domain. For identifying the performance metrics, on the basis of which to evaluate the SUT; existing OLTP, OLAP and HTAP relational database benchmarks and OLTP and OLAP graph database benchmarks will be consulted. Some HTAP relational database performance
metrics like throughput frontier and freshness as described in Milkai et al. [5] have already been identified which are suitable for use in evaluating HTAP graph database systems as they measure how the system balances the OLTP and OLAP workloads present in the HTAP workload.

After the problem domain is understood and the performance metrics have been identified, the benchmark components listed in section 3.1.1 will be designed and implemented.

4.2 Evaluation

After the benchmarking framework has been designed and implemented, the performance of HTAP graph database systems or related systems available for benchmarking will be evaluated. There are certain risks and challenges associated with this phase of the project, discussed in section 5.1, and therefore, the detailed evaluation methodology has been discussed in section 5.1.2 of this document.

5 Risks, Challenges, and Mitigation

This section describes some risks and challenges associated with the project.

5.1 Risk/Challenge 1: Absence of HTAP Graph Database Systems for Evaluation

5.1.1 Description

Since HTAP Graph Database Systems is a relatively new field of research, no commercial or academic implementations of systems of this type have been identified. There is a risk that no system of this type will be available for the evaluation phase of this project.

5.1.2 Proposed Mitigation

In the case where HTAP graph database systems are not available for benchmarking, recently proposed graph storage systems for HTAP graph database systems like LiveGraph proposed by Zhu et al. [10] and Sortledion proposed by Fuchs et al. [11] will be evaluated. This will be done by designing and implementing purpose-built databases for the benchmark schema using both storage systems and then executing the benchmark on them to compare their performance.
5.2 Risk/Challenge 2: Absence of Publicly Available Analytics Algorithms for Online Social Trading

5.1.1 Description

No publicly available analytical algorithms have been found for the online social trading domain so far. There is a risk that no such domain-specific analytical algorithms will be available for use in the benchmark.

5.1.2 Proposed Mitigation

In the case when such domain-specific algorithms are not available for use in the benchmark, the analytical algorithms in the LDBC Graphalytics benchmark proposed by Iosup et al. [12] will be used as substitutes. Iosup et al. [12] have presented the selection process in their paper, using which, they selected the algorithms to include in the LDBC Graphalytics benchmark and argue that these algorithms cover most usage scenarios. The LDBC Graphalytics benchmark has six algorithms: Breadth-First Search (BFS), PageRank (PR), Weakly Connected Components (WCC), Community Detection using Label Propagation (CDLP), Local Clustering Coefficient (LCC), and Single-Source Shortest Paths (SSSP) [12].

6 Project Milestones

This section lists the project milestones and the expected number of hours for each of them.

<table>
<thead>
<tr>
<th>Milestone</th>
<th>Expected Number of Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base Data Generator (Design and Implementation)</td>
<td>70</td>
</tr>
<tr>
<td>Workload Generator (Design and Implementation)</td>
<td>70</td>
</tr>
<tr>
<td>Benchmark Drivers (Design and Implementation)</td>
<td>60</td>
</tr>
<tr>
<td>Post-Benchmark Processor (Design and Implementation)</td>
<td>50</td>
</tr>
<tr>
<td>Evaluation</td>
<td>70</td>
</tr>
</tbody>
</table>

7 Project Schedule

This section defines the schedule for this project.
<table>
<thead>
<tr>
<th>Month</th>
<th>Tasks</th>
</tr>
</thead>
</table>
| September 2022  | - Literature Review  
                 - Project Plan  
                 - Project Webpage                                               |
| October 2022    | - Literature Review  
                 - Finalise Data Schema  
                 - Finalise Workload Design                                        |
| November 2022   | - Finish Base Data Generator Design  
                 - Start Base Data Generator Implementation                          |
| December 2022   | - Complete Base Data Generator Implementation  
                 - Test Base Data Generator  
                 - Generate Base Data for Evaluation  
                 - Design and Implement Workload Generator  
                 - Prepare for First Presentation  
                 - Prepare Detailed Interim Report                                    |
| January 2023    | - Test Workload Generator  
                 - Generate Workload for Evaluation  
                 - Design Query Drivers  
                 - Submit Detailed Interim Report                                     |
| February 2023   | - Implement Query Drivers  
                 - Test Query Drivers  
                 - Design and Implement Post-Benchmark Processor  
                 - Test Post-Benchmark Processor  
                 - Start Evaluation                                                    |
| March 2023      | - Evaluation  
                 - Final Report Preparation  
                 - Final Presentation Preparation                                      |
| April 2023      | - Completion of Deliverables  
                 - Submission of Deliverables                                           |

8 Conclusion

This document provided details on the plan and schedule to execute the project. The goal of the project is to develop a benchmarking framework for HTAP graph database systems and to
target the online social trading problem domain. Some challenges associated with this project and the plans to mitigate them were also discussed. It is hoped that following the proposed project plan will lead to successful execution of the project.

The project website is: https://wp.cs.hku.hk/2022/fyp22013.
9 References


