MSBD5014B Answering SQL Queries Under Differential Privacy Report

Name: MA, Zixin ID:21001887

**Introduction**

This project focused on implementing algorithms for achieving differential privacy (DP) in the context of the TPC-H benchmark queries. The goal of this project is to apply the concepts and techniques presented in two research papers: "R2T: Instance-optimal Truncation for Differentially Private Query Evaluation with Foreign Keys" [2] and "Shifted Inverse: A General Mechanism for Monotonic Functions under User Differential Privacy" [3].

**Background of the project and related work**

**Differential privacy** is a rigorous privacy framework that ensures the protection of individual data while allowing valuable analysis and querying of datasets. It guarantees that the results of queries on a dataset remain statistically indistinguishable even when an individual's data is included or excluded. By applying DP techniques, we can strike a balance between data utility and privacy, making it suitable for various domains, including SQL-based systems.

**TPC-H** is a widely used decision support benchmark that consists of a set of complex queries designed to evaluate the performance of database systems. By implementing DP algorithms for TPC-H queries, we can gain insights into the practical application of DP in real-world scenarios.

**Description of the query and the algorithm**

**Implementation details**

**Experimental setup and data sets used**

- Dataset Used

I generated dataset using TPCH benchmark[4] with 6 different scales: 0.125, 0.25, 1, 2, 4 which are marked as \_0.125, 0.25, 1, 2, 4, and store them in /data folder.

Command used: cd/TPCH/dbgen, ./dbgen -s <0.125>

- Import Data to Database

I used PostgreSQL database,

**Experimental results and discussion**

**Future directions**

**References**

[1]The Algorithmic Foundations of Differential Privacy, <https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf>

[2]R2T: Instance-optimal Truncation for Differentially Private Query Evaluation with Foreign Keys, <https://cse.hkust.edu.hk/~yike/R2T.pdf>

[3] Shifted Inverse: A General Mechanism for Monotonic Functions under User Differential Privacy, <https://cse.hkust.edu.hk/~yike/ShiftedInverse.pdf>

[4] <https://www.tpc.org/tpc_documents_current_versions/current_specifications5.asp>

[5]TPC-H表结构与22条查询SQL语句介绍, <https://segmentfault.com/a/1190000041322994>

[6]https://www.cnblogs.com/dajianshi/archive/2012/06/06/2827093.html

https://gist.github.com/yunpengn/6220ffc1b69cee5c861d93754e759d08