CSIT6910 Independent Project Proposal

Name: Ye Huaxiang
Student ID: 21187411
Program: MSc Information Technology, Fall 2025
Project Title: Instance-Optimal DP for SUM Queries over TPC-H – Implementing R2T with PostgreSQL & Python

1. Background & Motivation

TPC-H Q3 is a canonical 3-table join analytical query. We remove its GROUP BY clause and aim to release the total revenue under ε-differential privacy. The R2T algorithm achieves near-instance-optimal error for foreign-key scenarios and can be implemented on top of any RDBMS + LP solver.

2. Target Query (≥3 tables, single aggregate)

3. Objectives

Implement full R2T pipeline: join → LP construction → Laplace noise → max selection.
Verify ε-DP; measure accuracy vs PostgreSQL snapshot every 10 % of updates.
Performance: ≥10 k updates/s on 1 GB TPC-H (scale=1) under 1–8 cores.

4. Timeline (12 weeks)

Week	Milestone
1	Env setup: PostgreSQL, dbgen, Python, OR-Tools
2	Minimal R2T prototype on 1-table count
3–4	Extend to 3-table SUM query; correct LP output
5	Accuracy test vs PostgreSQL; error ≤5 %
6	Throughput benchmark (1 M updates, 1 core)
7	Parallel speed-up 1→2→4→8 cores
8	Vary $\varepsilon=\{0.1,0.5,1,2\}$; plot error-$\varepsilon$ curve
9	Baseline comparison: naive Laplace & smooth-sensitivity
10	Optional: MAX query via ShiftedInverse
11	Write 4–8-page report (figures + discussion)
12	Record 3-min demo & release GitHub repo

5. Deliverables

GitHub: code + README + notebook
Report PDF (4–8 pages)
Demo video + slides

6. Supervision

4 scheduled meetings (Wk 2/5/8/11); Ad-hoc questions via email to Jianzhuo or Prof. Yi