Concordia University
Dept. of Computer Science & Software Engineering
COMP 6521: Advanced Database Technology and Applications
Winter 2023

Lab Mini-Project 2

- Maximum Points: 9
- Reports: Submission via Moodle is due by Tuesday April 11th
- Demos: Wednesday April 12 during the Lab sessions

Description: We have a dataset of 10 million points in the 3D space stored in relation Points(X,Y,Z), where the attributes are real number in the range [0, 1000]. In this project you and your team are asked to develop an effective index structure (standard or ad hoc, tree based or hash based) to speed up finding answer(s) to the following TWO types of queries:

Q1 = {(X,Y,Z)| $x1 \leq X \leq x2$, $y1 \leq Y \leq y2$, $z1 \leq Z \leq z3$},

where x1, x2, y1, y2, z1, and z2 are some real numbers in the range [0, 1000]. Basically, Q1 asks to find all the points in the input dataset that is inside or lies on the borders of the cube defined by the query parameters x_i's, y_i's and z_i's. For instance, if x1=y1=z1=0 and x2=y2=z2=1000, then evaluation of Q1 will return all the 10,000,000 points in the input dataset. In order to help evaluate correctness of your implementation, report or display the number of answer tuples returned, and also store them in a file, called Q1-Output.

The second type of query Q2 would be as follows. Given any point A(x1, y1, z1) in the 3D space, find in the dataset the nearest neighbor point(s) of A. That is, the query would return EVERY point(s) in the relation Points that is closer to A. Note that the result may not be unique, in which case Q2 should return all such points that are at equal distance from A.

For both types of queries Q1 and Q2, you should be using the same index/hash structure you will be developing. In your project report, present the index/hash structure or mixed (ad hoc) and explain its particular features, if any. Also report the size of the index created. In your implementation and during your project demos, report/display the query processing time to answer these queries.

The maximum point to answer queries of type Q1 is 3 and of type Q2 is 3. The maximum point for the report (structure, content, and analysis) is 2. Presentation style and Q/A during your demo will get 1 point, at most.