Concordia University

Dept. of Computer Science and Software Engineering

COMP 6521: Advanced Database Technology and Applications

Winter 2019

Lab Assignment 2

>> Maximum Point: 8

>> Project Report: Wednesday April 8th

>> Project Demo: Wednesday April 10

Description: We have a large collection f tuples stored in relation Points(X,Y,Z). This relation has 10^7 tuples, each of which is a point in the 3D space and each coordinate is a real number in the range [0,1000], that is, $0 \leq X,Y,Z \leq 1000$. We will provide this relation to the groups. You and your team are required to develop an effective index structure (standard or ad hoc, tree based or hash based) to speed up answering 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 real numbers in [0,1000]. Basically, Q1 asks to find every point in the dataset of points that is inside or over the cube structure defined by the query. For instance if x1=y1=z1=0 and x2=y2=z2=1000, then the query result includes all the 10,000,000 points stored in the input dataset. In order to help evaluate correctness of your implementation, report or display the number of answer tuples returned, and if prompted by the user to “show the result”, then and only then display the result tuples or store the result in a file, say Q1.

The second type of query Q2 you should be able to answer against the above dataset of points is that, given any particular point A(x1,y1,z1), return the nearest neighbor(s) of A in the dataset, that is, the query result includes every point(s) in Points for which there is not closer point to A. The textbook and slides provide precise definition of nearest neighbor queries. For both types of queries Q1 and Q2, you should be using the same index structure you designed and developed. In your project report, present the index structure and explain any particular features it enjoys. Also report the size of the index

created. In your implementation and during your project demos, display the query processing time.

The maximum point to answer queries of type Q1 is 3 and of type Q2 is 2. The maximum point for report (structure, content, and analysis) is 2.
Good demos will get 1 point, at most.