INFS4205/7205 – Advanced Techniques for High Dimensional Data
Semester 1, 2021

Assignment 1 [*Total: 20 marks*]

Due date: 21 April 2021, 4:00 PM(AEST)
UQ Blackboard online submission only.

Consider three types of data in Queensland: one is about flora and fauna sightings which record the time and location (i.e., point) for a sighting of some species, one is about the areas of all defined wetlands (as polygons), and another is about the defined areas (as polygons) for state and national forest parks. A point is recorded using latitude and longitude, and a polygon is defined as a closed sequence of points.

Question 1 [*5 marks*] Design a database for these three types of data. You may choose to use spatial data types such as POINT, LINE and POLYGON. The tables you design should support at least the queries in Question 2 below. Use SQL create-table statements to present your design [*You can document any your design assumptions if necessary*].

Question 2 [*5 marks*] Write three queries in an SQL-like query language,
  (1) [*1 marks*] to find the number of sightings of legless lizards in Pine Ridge Conservation Park.
  (2) [*2 marks*] to find all wetlands inside a state or national forest park.
  (3) [*2 marks*] to find all sightings of platypus and the distance to the closest wetlands (set the distance to 0 if the sighting is inside a wetland).
[*Please define any non-standard SQL operations, relationships and functions you choose to use in a query.*]

Question 3 [*10 marks*] Design a step-by-step query execution plan, aiming to minimize the data to be fetched from the database and the number of spatial operations to be performed.
  (1) [*3 marks*] to process the query you give in Question 2.1 using Quad-tree indexes.
  (2) [*3 marks*] to process the query you give in Question 2.2 using R-tree indexes.
  (3) [*4 marks*] to process the query you give in Question 2.3 using R-tree indexes.
[*All data structures including spatial indexes used in the execution plan need to be clearly described. Execution plan description can be in either plain English, or some kind of pseudo code, and the description must be clear and concise. You can make use of non-spatial indexes too if it is beneficial. Pleases explain how the cost of data access can be reduced by using spatial indexes and your execution plan.*]