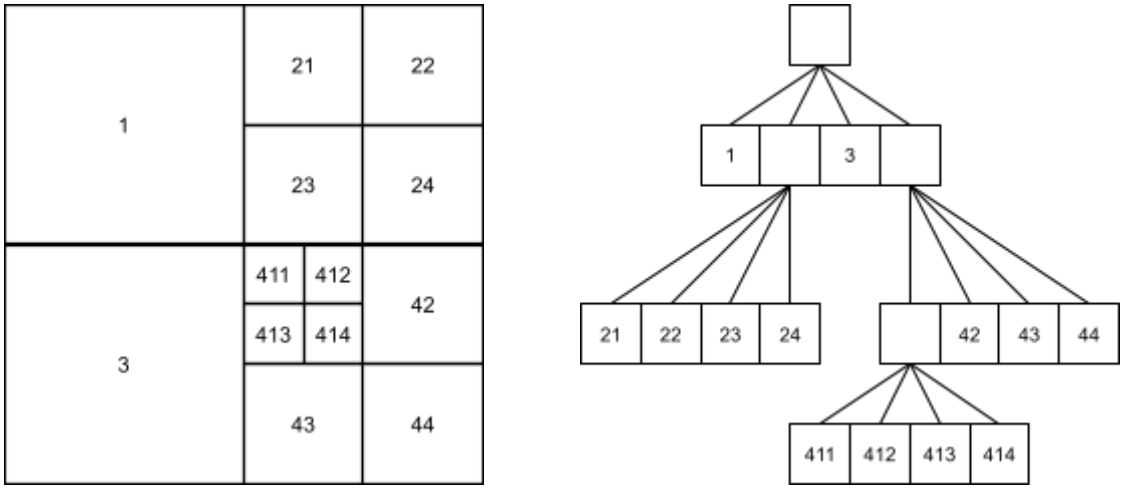
CAS CS 561: Data Systems Architectures
Data-intensive Systems and Computing Lab
Department of Computer Science
College of Arts and Sciences, Boston University
bu-disc.github.io/CS561

Robust Ingestion for Spatial Indexes

Mentor: Konstantinos Karatsenidis

Background

Spatial indexes, such as R-trees and Quadtrees, are data structures designed to efficiently organize and retrieve spatial data. R-trees use bounding boxes to define regions, while Quadtrees partition space into quadrants. These indexes are optimized for spatial queries for applications like geographic information systems and location-based services, particularly in scenarios where consecutive data points exhibit spatial locality, as seen in GPS tracking.



A simple example of a Quadtree

In streaming scenarios, like real-time GPS tracking of a moving vehicle, spatial locality is evident as consecutive data points closely follow the vehicle's route. Existing implementations of spatial indexes overlook this spatial coherence and process each point independently, missing the opportunity to optimize computations based on the predictable movement pattern. By harnessing the spatial relationships between these streaming data points, a more efficient system could ingest data faster, minimize computational redundancies, and improve the responsiveness of the database system.

Objective

The goal of this project is to implement a spatial index that keeps track of the latest inserts. This index should be able to keep multiple active nodes as the application may require tracking multiple tracking devices concurrently. You will also have to brainstorm ways to predict where the tracking device is heading so that the active node can follow this path.