Scalable Distributed Processing of K Nearest Neighbor Queries over Moving Objects

Abstract

Central to many applications involving moving objects is the task of processing k-nearest neighbor (k-NN) queries. Most of the existing approaches to this problem are designed for the centralized setting where query processing takes place on a single server; it is difficult, if not impossible, for them to scale to a distributed setting to handle the vast volume of data and concurrent queries that are increasingly common in those applications. To address this problem, we propose a suite of solutions that can support scalable distributed processing of k-NN queries. We first present a new index structure called Dynamic Strip Index (DSI), which can better adapt to different data distributions than existing grid indexes. Moreover, it can be naturally distributed across the cluster, therefore lending itself well to distributed processing. We further propose a distributed k-NN search (DKNN) algorithm based on DSI. DKNN avoids having an uncertain number of potentially expensive iterations, and is thus more efficient and more predictable than existing approaches. DSI and DKNN are implemented on Apache S4, an open-source platform for distributed stream processing. We perform extensive experiments to study the characteristics of DSI and DKNN, and compare them with three baseline methods. Experimental results show that our proposal scales well and significantly outperforms the alternative methods.
EXISTING SYSTEM

Moving objects is the task of processing k-nearest neighbor (k-NN) queries. Single server can be processing the query to scale to a distributed setting to handle the vast volume of data and concurrent queries that are increasingly common in those applications.

DRAWBACK OF EXISTING SYSTEM

- The volume of data in the Internet scale, because they implicitly assume a centralized setting.
- Light update.
- Query workload.
- No longer viable.
- Exceeds the capacity of a single server.

PROPOSED SYSTEM

Scalable distributed processing of k-NN query method used in proposed system Dynamic Strip Index (DSI), adapt to different data distributions than exiting grid indexes. Distributed cluster that consists of a single master and multiple slaves. More predictable performance than existing grid-based approaches.

ADVANTAGE OF PROPOSED SYSTEM

- More efficient and more predictable
- Updates can be handled quickly.
• Good load balancing

**SYSTEM SPECIFICATION**

**Hardware Requirements**

- System: Pentium IV 2.4 GHz
- Hard Disk: 40 GB
- Floppy Drive: 1.44 Mb
- Monitor: 15 VGA Colour
- Mouse: Logitech
- Ram: 512 Mb

**Software Requirements**

- Operating system: Windows Family
- Tools: eclipse
- Technology Used: Java
- Backend Used: SQLITE