Turkish Journal of Computer and Mathematics Education *Vol.12 No.1S (2021), 106-114*

Research Article

Location based Continuous Query Processing over Geo-streaming Data

K. V. Metre[1], M. U. Kharat[2]

[1]MET's Institute of Engineering, Nashik, India
[2]MET's Institute of Engineering, Nashik, India

Article History: Received: 10 November 2020; Revised: 12 January 2021; Accepted: 27 January 2021; Published online: 05 April 2021

Abstract: In recent years, many data-intensive and location based applications have emerged that need to process stream data in applications such as network monitoring, telecommunications data management, and sensor networks. Unlike regular queries, a continuous query exists for certain period of time and need to be continuously processed during this time. The algorithms used for data processing for the traditional database systems are not suited to tackle complex and various continuous queries over dynamic streaming data. The indexing for finite queries is preferred to indexing on infinite data to avoid expensive operations of index maintenance. Previous related work focused on moving queries on static objects or static queries on moving object. But now-a-days queries as well as objects are dynamic. So, hybrid indexing for queries significantly reduces the space costs and scales well with the increasing data. To deal with the speed of unbounded data, it is necessary to use data parallelism in query processing. The data parallelism in query processing offers better performance, availability and scalability.

Keywords: Data parallelism, hash value, space filling curve, spatial-temporal query

1. Introduction

In recent years, new data-intensive applications, such as network monitoring, traffic monitoring, sensor networks, telecommunications data management, and others, involve data streams. The processing of high volume data stream has cropped up as a research issue in various engineering and scientific applications e.g. monitoring mobile objects (Wang, 2011) tracking continuous network behaviors (Bohm, 2007) processing of data from sensors (Carney, 2002) and identification of RFID data (Park, 2007). To filter the data which is rapidly coming at the system as time-varying, unbounded sequences of data objects, the user uses continuous queries as against one-time submitted queries. These queries are processed over streams continuously for a period of time and provide new results incrementally on the arrival of new data tuples. The timestamps or time intervals are considered for system-wide windows for each single processing step (Golab, 2003). The authors introduced the adaptive windowing technique which dynamically resizes the window based on the incoming data (Bifet , 2007). (Babu, 2001) discussed requirements and challenges in query processing, and algorithmic issues.

The continuous queries are submitted once and executed continuously to get the output from data streams (Golab, 2003). One of the biggest challenges for a Data Stream Management System is to process continuous and massive data streams in consideration of the memory constraints for unbounded data. Window techniques focus on recent portion of the data. The recent data can be taken in terms of n objects or the objects during a time span. Re-execution of queries is a costly operation, and processing of continuous queries is considered an open issue in data stream systems using various indexing approaches. The data indexing becomes expensive as the data is not persistent, but it is in large volume. So, a new approach is suggested to build the index for queries which is finite rather than to build the index for data which is infinite (Carney, 2002). To evaluate location-based queries, the indexing for data or queries plays a key role. The various approaches are used for indexing spatial queries and spatial data which are categorized into (a) Tree-based structures (b) Grid structures and (c) Space-filing curves. Most applicable to the work presented here are space-filling curves and tree-based methods. The tree-based approaches such as R-tree (Prabhakar, 2002) incur a high updating cost due to tree structure. The grid-based methods use cell approach (Bohm, 2007; Kalashnikov, 2004) for indexing spatial queries and it divides the indexing pace into equal-sized grid. The grid-based approach results in better search and update time than the tree-based approaches (Kalashnikov, 2004) but it incurs high construction cost for indexing queries (Chen, 2011). Now days, due to dynamic nature of location-based queries, various query indexing approaches for continuous range queries (Gedik, 2006 ; Li & Karras, 2012) are suggested. Query Indexing mainly is suitable for evaluating continuous queries over dynamic and moving objects, as the queries exist for long periods of time, whereas the objects are continuously moving (Kalashnikov , 2002).

As the data is coming continuous and in unbounded manner, so it becomes a big challenge to filter massive data stream efficiently using query index. The query indexing method must support (1) scalability to massive data stream and (2) continuous queries. As we filter massive data which is coming continuously and the relevant objects for the given query are returned, it is rational to relinquish optimality with a good feasible solution that can be computed efficiently. The approximation methods provide reasonable and feasible solution instead of exact optimal solution.