Real-Time Data Stream Partitioning over a Sliding Window in Real-Time Spatial Big Data

Authors : Sana Hamdi, Emna Bouazizi, Sami Faiz

Abstract : In recent years, real-time spatial applications, like location-aware services and traffic monitoring, have become more and more important. Such applications result dynamic environments where data as well as queries are continuously moving. As a result, there is a tremendous amount of real-time spatial data generated every day. The growth of the data volume seems to outspeed the advance of our computing infrastructure. For instance, in real-time spatial Big Data, users expect to receive the results of each query within a short time period without holding in account the load of the system. But with a huge amount of real-time spatial data generated, the system performance degrades rapidly especially in overload situations. To solve this problem, we propose the use of data partitioning as an optimization technique. Traditional horizontal and vertical partitioning can increase the performance of the system and simplify data management. But they remain insufficient for real-time spatial Big data; they can’t deal with real-time and stream queries efficiently. Thus, in this paper, we propose a novel data partitioning approach for real-time spatial Big data named VPA-RTSBD (Vertical Partitioning Approach for Real-Time Spatial Big data). This contribution is an implementation of the Matching algorithm for traditional vertical partitioning. We find, firstly, the optimal attribute sequence by the use of Matching algorithm. Then, we propose a new cost model used for database partitioning, for keeping the data amount of each partition more balanced limit and for providing a parallel execution guarantees for the most frequent queries. VPA-RTSBD aims to obtain a real-time partitioning scheme and deals with stream data. It improves the performance of query execution by maximizing the degree of parallel execution. This affects QoS (Quality Of Service) improvement in real-time spatial Big Data especially with a huge volume of stream data. The performance of our contribution is evaluated via simulation experiments. The results show that the proposed algorithm is both efficient and scalable, and that it outperforms comparable algorithms.

Keywords : real-time spatial big data, quality of service, vertical partitioning, horizontal partitioning, matching algorithm, hamming distance, stream query