Information Systems 44 (2014) 134–154

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/infosys

Parallel online spatial and temporal aggregations on multi-core CPUs and many-core GPUs

Jianting Zhang [a,*], Simin You [b], Le Gruenwald [c]

[a] Department of Computer Science, the City College of New York, New York, NY 10031, USA
[b] Department of Computer Science, CUNY Graduate Center, New York, NY 10016, USA
[c] School of Computer Science, the University of Oklahoma, Norman, OK 73071, USA

ARTICLE INFO

Available online 22 January 2014

Keywords:
OLAP
Parallel design
GPU
Multi-core CPU
Spatiotemporal aggregation
Spatial indexing
Spatial join
Large-scale data

ABSTRACT

With the increasing availability of locating and navigation technologies on portable wireless devices, huge amounts of location data are being captured at ever growing rates. Spatial and temporal aggregations in an Online Analytical Processing (OLAP) setting for the large-scale ubiquitous urban sensing data play an important role in understanding urban dynamics and facilitating decision making. Unfortunately, existing spatial, temporal and spatiotemporal OLAP techniques are mostly based on traditional computing frameworks, i.e., disk-resident systems on uniprocessors based on serial algorithms, which makes them incapable of handling large-scale data on parallel hardware architectures that have already been equipped with commodity computers. In this study, we report our designs, implementations and experiments on developing a data management platform and a set of parallel techniques to support high-performance online spatial and temporal aggregations on multi-core CPUs and many-core Graphics Processing Units (GPUs). Our experiment results show that we are able to spatially associate nearly 170 million taxi pickup location points with their nearest street segments among 147,011 candidates in about 5–25 s on both an Nvidia Quadro 6000 GPU device and dual Intel Xeon E5405 quad-core CPUs when their Vector Processing Units (VPUs) are utilized for computing intensive tasks. After spatially associating points with road segments, spatial, temporal and spatiotemporal aggregations are reduced to relational aggregations and can be processed in the order of a fraction of a second on both GPUs and multi-core CPUs. In addition to demonstrating the feasibility of building a high-performance OLAP system for processing large-scale taxi trip data for real-time, interactive data explorations, our work also opens the paths to achieving even higher OLAP query efficiency for large-scale applications through integrating domain-specific data management platforms, novel parallel data structures and algorithm designs, and hardware architecture friendly implementations.

1. Introduction

With the increasing availability of locating and navigation technologies on portable wireless devices, huge amounts of location data are being captured at ever growing rates. For example, the approximately 13,000 taxicabs in the New York City (NYC) equipped with GPS devices generate more than half a million taxi trip records per day. Cell phone call logs represent a category of data at an even larger scale [1,2]. Also as visitors travel around the world more frequently, location-dependent social networks such as Foursquare [3], and location-enhanced social media such as text posted to Wiki sites [4], and images and videos posted to Flickr and

* Corresponding author. Tel.: +1 212 650 6175.
E-mail addresses: jzhang@cs.ccny.cuny.edu (J. Zhang), syou@gc.cuny.edu (S. You), ggruenwald@ou.edu (L. Gruenwald).

http://dx.doi.org/10.1016/j.is.2014.01.005
0306-4379

YouTube [5], can also potentially generate large-amounts of spatial and temporal data. All the three types of data have a few features in common: (1) they are produced and collected by commodity sensing devices and are rich in data volumes in urban areas; and (2) they are a special type of spatial and temporal data with an origin location and a destination location in the geo-referenced space domain and a starting time and an ending time in the time domain. However, the intermediate locations between origins and destinations are either unavailable, inaccessible or unimportant. Compared with traditional geographical data collected by government agencies for urban planning and city administration purposes, these data can be more effective to help people understand the real dynamic of urban areas with respect to spatial/temporal resolutions and representativeness. We term such data as Ubiquitous Urban Sensing Origin-Destination data, or U^2SOD data, for notation convenience [6]. Despite the close relationships between U^2SOD data and Spatial Databases (SDBs) [7] and Moving Object Databases (MODs) [8], our experiences have shown that traditional disk-resident and tuple/row oriented spatial databases and moving object databases are ineffective in processing large-scale U^2SOD data for practical applications including multi-dimensional aggregations, one of the most important modules in Online Analytical Processing (OLAP) [9].

Considerable work on developing efficient data structures and algorithms has been proposed for multidimensional aggregations on CPU uniprocessors in the past few decades [10]. Modern hardware architectures increasingly rely on parallel technologies to increase the processing power due to various limits in improving the speeds of uniprocessors [11]. Unfortunately, existing data structures and algorithms that are designed for serial implementations may not be able to effectively utilize the parallel processing power of modern hardware, including multi-core CPUs and many-core Graphics Processing Units (GPUs) [11]. Despite the fact that parallel hardware is already available in the majority of commodity computers, there is still relatively little work in exploiting such parallel processing power for OLAP queries, especially in the areas of spatial and temporal aggregations of large-scale geographical data where complex join operations are required in the aggregations. Examples of these are counting the number of taxi pickups at each of the community districts or census blocks (spatial aggregation), generating hourly histogram of drop-offs near the JFK airport (temporal aggregation) and computing numbers of trips between Time Square and Central Park in morning peak hours (OD aggregation).

In this study, we report our work on developing a data management framework for large-scale U^2SOD data and a set of data parallel designs of spatial and temporal aggregations that can be realized on both multi-core CPUs and many-core GPUs in an OLAP setting. Our experiments on aggregating the approximately 170 million taxi trip records in NYC in 2009 have demonstrated the effectiveness of the proposed framework and parallel techniques. By utilizing parallel spatial joins [12] to support efficient online processing, we are able to achieve real-time responses for spatial, temporal and spatiotemporal aggregations at different hierarchical levels. Compared with traditional approaches that rely on relational databases and spatial databases for aggregations, our techniques have reduced the OLAP query response times from hours to seconds. This makes it possible for urban geographers and transportation researchers to explore the large-scale origin–destination data in general and taxi trip data in particular in an interactive manner.

The rest of the paper is arranged as follows. Section 2 introduces the background, motivation and related work. Section 3 presents a parallelization-friendly data management framework for managing large-scale origin–destination data with a focus on taxi trip records. Section 4 provides the details on the parallel designs and implementations of spatial and temporal aggregations on both multi-core CPUs and many-core GPUs. Section 5 reports the experiment results. Finally Section 6 is the conclusion and future work.

2. Background, motivation and related work

The increasingly available location data generated by consumer wireless portable devices, such as GPS, GPS enhanced cameras and GPS/WiFi/Cellular enhanced mobile phones, has significantly changed the ways of collecting, analyzing, disseminating and utilizing urban sensing data. Traditionally city government agencies are responsible for collecting various types of geographical data for city management purposes, such as urban planning and traffic control. The data collection is usually done through sampling, typically at coarse-resolutions, and questionnaire-based investigations which often incur long turn-around times. In contrast, as consumer mobile devices become ubiquitous, similar data obtained from GPS-traces and mobile phone call logs has much finer resolutions. With the help of privacy and security related technologies, the aggregated records from such ubiquitous urban sensing data can be enormously helpful in understanding and addressing a variety of urban related issues. Research groups from both academia (e.g., MIT Sensible City Lab) and industries (e.g., IBM Smart Planet Initiative and Microsoft Research Asia) have developed techniques to utilize such data and understand the interactions among people and their locations/mobility at the city level (e.g., Beijing [13], Boston [14] and Rome [1]), social group level (e.g., friends [15] and taxi-passenger pairs [16]) and individuals level [17]. However, most of the existing studies focus on the data mining aspects of such ubiquitous urban sensing data through case studies while largely leaving the data management aspects untouched. Lacking proper data management techniques can result in significant technical hurdles in making full use of such data to address outstanding societal concerns. In this study, we focus on efficiently aggregating large-scale taxi trip records to better understand human mobility and facilitate transportation planning by developing high-performance spatial, temporal and spatiotemporal aggregation techniques in an OLAP setting.

OLAP technologies are attractive to explore the possible patterns from large-scale taxi trip records and other types of origin–destination data. As the taxi trip data has spatial dimensions and temporal dimensions for both pickup and