A Real-Time System for Air Quality Monitoring Based on Main-Memory Database

Zhuo Tang[1], Wei Xiong[2]*, Luo Chen[2], Ning Jing[2]

[1] Department of Geriatrics, Hunan Provincial People's Hospital, Changsha, China

[2] Department of Information and Communication Engineering,
College of Electronic Science and Engineering, National University of Defense Technology, Changsha, China

*Corresponding author, e-mail: xiongwei@nudt.edu.cn

Abstract—**Air pollution problems are closely linked to geographic locations. The existing environmental monitoring systems have little focus on geographic information processing and real time visualization. In this paper, a spatial extension to main-memory database is implemented to support the storage and index of spatial-temporal data. Meanwhile, historic data is stored in disk-based spatial database, providing the function of querying and analyzing. A real-time air quality monitoring system is designed and implemented which can process and display data flow on real-time. Moreover, it can deal with concurrent queries from multi-users effectively. The visualization module is cross-platform, supporting both 2-dimensional and 3-dimensional visualization.**

Keywords-main-memory database; air quality; real-time; spatial index; 3-dimensional;

I. INTRODUCTION

Due to the irrational of energy structure and people's weak environmental awareness, the air environmental pollution problem is getting more serious and has caused grave impact on economic development and people's life. Nowadays, the spread of internet and intelligent mobile devices has provided convenience to public to know the air quality in time. Nevertheless, current systems for environment monitoring are rarely integrated with real-time geographical information efficiently and visualized in rich ways [1]. Instead of the way by using a certain numerical value to stand for the air quality of a whole region, the users prefer to watch the distribution situation and the changes over time of air quality in a map. Besides, they are happy to view air quality both in two-dimensional and three-dimensional (3D) ways and retrieve and check the air quality condition of a certain time in the past. It is difficult for efficiently analysis and visualizing based on traditional air quality data organization methods. If a system can store, analyze and display air quality data in real-time, it will be an aid decision making tool for all levels of environmental departments. Whereas little existing environmental monitoring systems present satisfied performance in geographic data processing and visualization when facing massive dataset, especially for real-time data [2].

With the rapid development of wireless network and communication technology, there is more feasible that air quality monitoring system is designed in wireless mode. Wireless Sensor Network (WSN) has been rapidly developed during recent years, which brings great challenges to rapid processing for air quality data streams. Traditional air quality monitoring systems based on WSN mainly focus on how to effectively manage energy in sensor network [3]. These systems can provide high-quality air quality measurements except for high concurrent internet service [2]. In real-time applications, air quality monitoring system need to provide high performance in calculation and processing of various air quality values [4]. For a decision support system, spatial-temporal analysis and mapping is very important. Some air quality monitoring systems apply geographical information system (GIS) to implement these functions [5]. However, the performance and diverse visualization is not well addressed.

A real-time system for air quality monitoring based on main-memory database under high performance hardware environment is designed and implemented. Goals of this system are as follows: (1) providing a hybrid storage mode with main-memory database and traditional spatial database to deal with continuous concurrent query, (2) supporting abundant and intuitive way of presenting, (3) playback of historic data.

II. SYSTEM ARCHITECTURE

This system is based on a high-performance geographic information system, HiGIS [7]. HiGIS can be used to improve the performance of time-consuming GIS operations by utilizing parallel computing in a HPC environment, and to build an open framework for geospatial data access, processing, analysis, and visualization. The backend of HiGIS is running on a HPC (High Performance Computing) cluster, while the frontend is fairly thin and cross-platform. The architecture of air quality monitoring system build on the HiGIS is showed as figure 1.

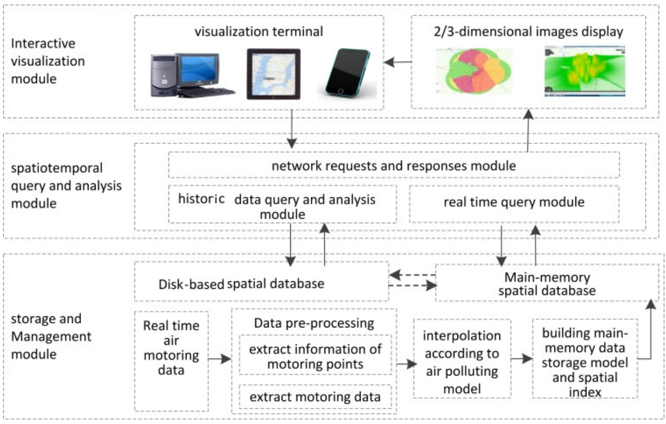


Figure 1. System architecture

Supported by the National High Technology Research and Development Program of China (Grant No. 2015AA123901.) and the National Natural Science Foundation of China (Grants No. 41271403 and Grant No. 41471321).

A. *Storage and management module*

Storage and management module is the basic module for data organization of this system. It receives the monitoring data from air quality monitoring sensors and processes and stores them in real time. Meanwhile, it provides query function for upper module. After received monitoring data, this module needs to extract names, coordinates and the monitoring values from the data stream of monitoring sensors. Due to the sparse distribution of the monitoring sensors, we need to use certain spatial interpolation model to numerical interpolate to intuitively present the distribution situation of the air pollution in the whole monitoring area. In order to meet the real time processing requirement, the interpolation is worked on high-performance Linux server.

In this system, the monitoring data of air quality monitoring sensors are transmitted to the server via data stream. After the interpolation, data quantity increased rapidly. Meanwhile, the data stream needs to be fast presented in real time in the client side. As traditional disk-based relational database cannot satisfy the performance requirements, we proposed a main-memory database with spatial extension to make the data transmit from server to client directly omitting the I/O progress on disk. As the capacity of main memory in the server is limited and it would loss data when power is off, the system is using a hybrid storage mode combining main-memory database and disk-based database. Among them, the main-memory spatial database is responsible for store, manage and analyze the "hot data" while the disk-based spatial database is in charge of data persistence and querying and analyzing historic data (i.e. cold data). Main-memory database and disk-based database synchronize data when the main memory usage reaches the upper limit or the system is failure.

B. *Spatiotemporal query and analysis module*

This module consists of three functions, network communication, real time query, and historic data query.

Network communication receives requests of spatial query and analysis from interactive visualization module, and translates the query requests to the spatial query plan. Then it sends the transformed query plan to corresponding disk-based or main-memory spatial database. After the spatial query executed, it returns the results to the client and prepares data for the visualization.

Real time query connects to main-memory spatial database and processes the requests of real-time data visualization. Concurrent users' real-time query requests, such as the current air quality distribution and the dynamic changing situation in a certain region, are also processed by this module.

The historic data query processes the query requests of historic data and the corresponding spatiotemporal analysis such as the query of checking the air quality changing of a certain region in a period of time, or the number of days in the past one month that air quality is worse than middle level pollution.

C. *Interactive visualization module*

Visualization module is made up of two parts that are two-dimensional module and three-dimensional module. The client side can be desktop, tablet computer, mobile phone and other mobile devices. Only a browser application is needed to be installed in the device.

Two-dimensional visualization is implemented based on HTML5 visualization engine. It presents the air quality condition of all the locations covered by the sensors. The area covered by the same sensor is rendered in the same color, and they are updated in real time according to the streaming data. Meanwhile, information about air quality monitoring sensors can be retrieved at any time.

Three-dimensional visualization is implemented based on WebGL. It allows the browser to invoke GPU directly without installing any plug-ins. It is different from just only using color to represent the air quality in 2D plane, that 3D visualization can display the air quality information in diversified means. 3D visualization uses colored cylinders for presenting more detailed air quality. The height of the cylinders is the actual air quality numerical value and the color stands for the air quality classification color. By this means, it is intuitively to watch the changing of air quality with time while the cylinders' height is changing.

III. KEY TECHNIQUES

There are two main key technologies involved in this system including spatial data storage and retrieve in the main-memory database and rapid 3D visualization.

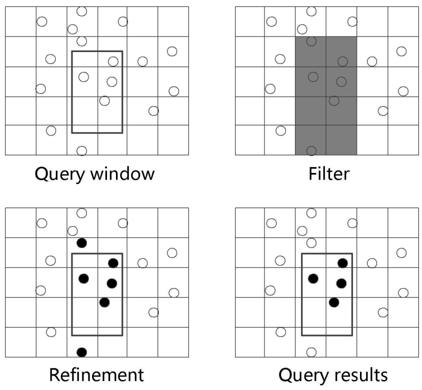


Figure 2. Spatial query progressing in main-memory database

Data management in main-memory database is quite different from traditional spatial database. The main difference is that all spatial data being reorganized and managed in the main memory, the performance of data access and processing increased significantly when comparing to disk-based relational database. Therefore, to fulfill the requirements of rapid responding and frequent concurrent reading and writing, main memory database is a perfect choice. Redis is an open source key-value storage system which stores the data in the main memory [8]. However, main memory database doesn't support spatial data model and index directly. If we want to use Redis, we have to extend its spatial data types and spatial index first [9]. Extending spatial types is implemented by modifying Redis's source codes. Adding two functions, setg and getg, to write and read the spatial data. Spatial data are stored in main memory as the form of WKB (Well-Known Binary), and presented as the form of WKT (Well-Known Text).

Likewise, grid index based on geohash is extended for indexing multi-dimensional spatial data. Grid index is well-suited for querying spatial points, and the search progress using grid index is showed as figure 2.

For line and polygon objects, traditional geohash encoding method has some problems. If using larger grid to represent spatial object, the objects with small size cannot be reflected in actual size. If using smaller grid to represent spatial objects, the objects with large size may occupy more grids, which is not efficient for spatial query. Therefore, an adaptive geohash encoding method is proposed for indexing the line and polygon objects efficiently. In this method, an object has 4 geohash codes at most, which is shown in figure 3.

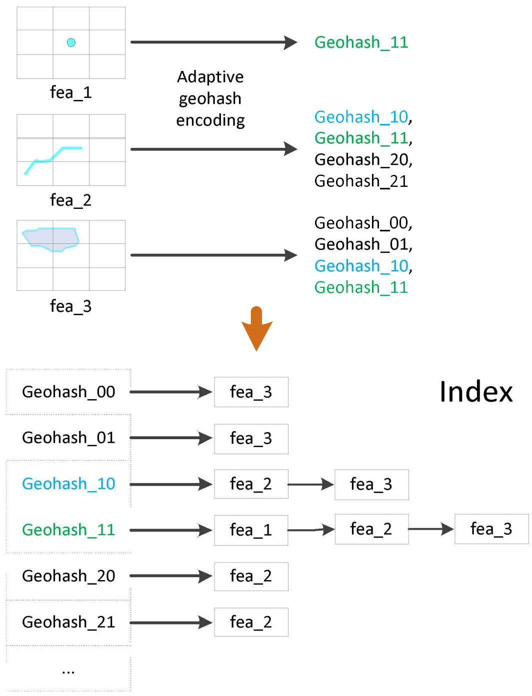
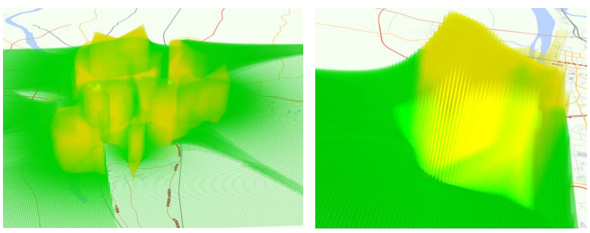


Figure 3. Adaptive geohash encoding for spatial index

3D visualization is more complicated than two-dimensional visualization, which requires utilizing graphic processing unit (GPU) directly. To ensure the fluency of 3D rendering, a strategy imitating the changes of human vision is proposed. 3D rendering are implemented according to different particle sizes zoomed from users. When zooming in, it will draw with fine-grain with delicate display. When zooming out, it will take samples from data to render keeping the display quantity not to be affected.



(a) coarse-grain display (b) fine-grain display

Figure 4. 3D rendering strategy with different particle size

IV. Experiments and System Prototype

The configurations of prototype system are as follows:

Server: SuperMicro Tower server, configuration: Intel® Xeon® E5-4620 CPU@2.20GHz, 500GB DDR3 RAM, 16TB storage array, CentOS 6.3 operating system, 10G Ethernet switches.

Client: (1)PC: Intel® CoreTM i5-4570 CPU@3.2GHz, 8GB DDR3 RAM, Windows7 operating system, Google Chrome 41.0 browser; (2) tablet computer: Apple A7 processor, 1GB RAM, IOS 8.2 operating system, Safari browser; (3) mobile phone: 4-core Qualcomm snapdragon 600 processor@1.7GHz, 2GB DDR3 RAM, Android 4.2.2 operating system, Chrome browser.

Dataset: 470 days real air quality monitoring data of Changsha from December 2nd, 2013 to March 16th, 2015 and some simulated real-time data. Location of monitoring points and value of AQI, PM2.5, PM10, SO_2, CO, NO_2 and O_3 and grading information are included in the data [6].

Firstly we evaluated the query performance of main-memory spatial database compared to PostGIS. In this experiment, we use the dataset as table 1. With the increasing of size of query window, it can be observed that point query performance is better than line query in main-memory spatial database. In geohash index, one line objects may corresponds to multiple geohash codes. Therefore, geohash index is more proper for point query. In air quality monitoring system, most used query type is point query.

Table 1. Dataset for Query Performance Evaluation

Dataset	Number of features
Beijing points	251282
Beijing roads	144162

Windows query performance changes with the size of query window. Point query and line query can be shown in figure 5 and figure 6 respectively.

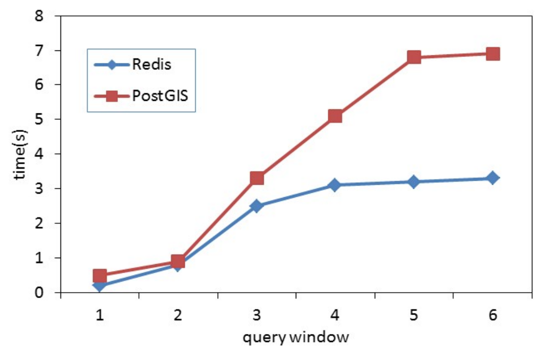


Figure 5. Point query performance comparison

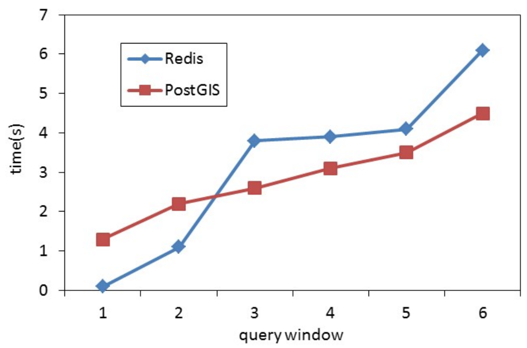


Figure 6. Line query performance comparison

In figure 7, the concurrent query performance comparison shows that main-memory spatial database have more fast average response time than traditional spatial database. That means when facing multiple users, main-memory database is more proper for air quality query in web environment.

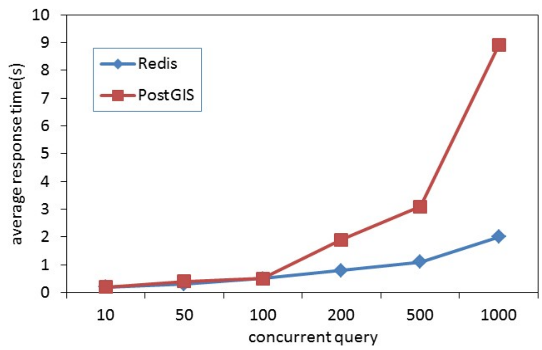


Figure 7. concurrent query performance comparison

This system has integrated in HiGIS, which is an open framework for high performance geocomputation [7]. The main goal of HiGIS is to improve the performance of processing by utilizing parallel computing techniques. The air quality of PM2.5 will be presented in real time as showed in figure 8. Different colors stand for the variety of degrees of pollution. The darker the color, the pollution condition is more serious. With the inflowing of monitoring data, the image is refreshing in real time.

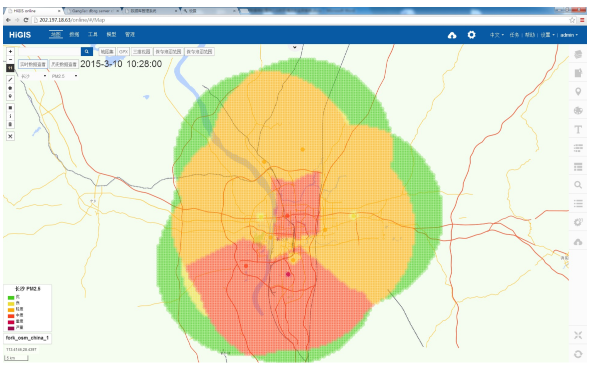


Figure 8. Real-time monitoring data of PM2.5 in Changsha

Although the efficiency of inquiry for these data is not a problem in a single user request, the system must be confronted test from high concurrent query requests when real-time data stream arriving. When clicking at the point of interest, the location, pollutant value and grade will be showing. Besides the real time motoring, this system supports retrieving and replaying of historic data. Two retrieve modes are supported: (1) setting the query date to get the air quality, (2) using a progress bar for a period to play the historic data. The 3D view of air quality is also presented in this system as showed in figure 9.

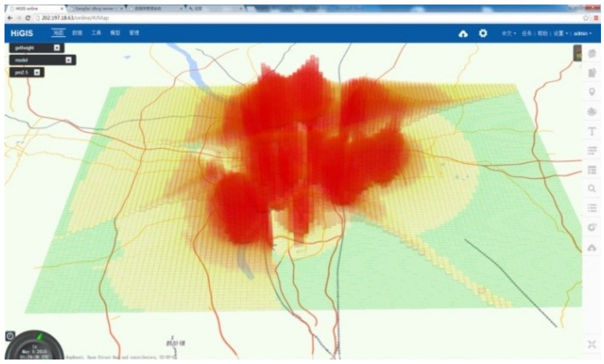


Figure 9. 3D view of air quality

V. Conclusion and Future Work

This prototype system relies on high-performance server. It can process and visualize the air quality monitoring data flow in real time. A hybrid storage method combining the main-memory database with traditional spatial database can efficiently support high concurrent real-time query and historic data query.

In the future, we will need to add more functions such as optimizing the appearance of visualization, optimizing interpolation algorithm, adding the statistics and estimation function of air quality, related analyzing historic data, mining implicit information, etc.

References

[1] S. Davila, B. Ivan, Š. Krešimir. "Information Systems for Air Quality Monitoring." Air Quality-New Perspective. InTech, 2012.

[2] H. Ali, J. K. Soe, S. Weller. "A real-time ambient air quality monitoring wireless sensor network for schools in smart cities." Smart Cities Conference (ISC2), 2015 IEEE First International. IEEE, 2015: 1-6.

[3] G. Swagarya, S. Kaijage, R. S. Sinde. "A Survey on Wireless Sensor Networks Application for Air Pollution Monitoring.". International Journal of Engineering And Computer Science, 2014, 3(5): 5975-5979.

[4] S. Raipure. "Calculating Pollution in Metropolitan Cities using Wireless Sensor Network". International Journal of Advance Research in Computer Science and Management Studies, 2014, 2(12): 293-296

[5] T. Elbir. "A GIS based decision support system for estimation, visualization and analysis of air pollution for large Turkish cities." Atmospheric Environment, 2004, 38(27): 4509-4517.

[6] http://www.pm25.in/changsha

[7] W Xiong, L Chen. "HiGIS: An Open Framework for High Performance Geographic Information System." Advances in Electrical and Computer Engineering 15.3 (2015): 123-132.

[8] http://www.redis.io/

[9] K. Alfons, T. Neumann. "Main-memory database systems." 2014 IEEE 30th International Conference on Data Engineering (ICDE). IEEE, 2014:1310.