Study on Graphic Database & Visual Query System of Ecological Development of Dianchi Lake Basin

Jie Song[1, a *], Xingyuan Chen[1,b]

[1]School of Information Technology, Kunming University, Kunming, China

[a]1353257038@qq.com, [b]95430601@qq.com

Abstract. It is rather difficult to store and manage the diverse data source of ecological development of Dianchi Lake basin through traditional RDBMS due to their various types, massive data, complex relationship as well as their dynamic changes. We established graphics database and visual query system to analyse the ecological system of Dianchi Lake basin based on targeted modeling and study of query methods on ecological environment characteristics, by adopting graph store models of Neo4j, a high-performance distributed graphics database, to solve the necessity to structure index and query of multiple data sources under distributed environment by the time of data renewal. You can rely on the traverse and search of the graphics to add or delete the database, which is less complex and has a relative faster speed for data renewal. It adopts composite algorithm of incremental subgraph searching to solve the unsatisfactory result of query due to context ambiguity from key word query, and brings a high matching accuracy. By adopting visual query, users could have access to a friendly graphics interface, and have access of feedback graphics for every step of incremental query. It changes the situation that the query result cannot be observed directly and reflects the mutual relationship between complex data and plays an important role on scientific decision making of ecological development.

Introduction

The following factors contribute to the degradation of ecological environment quality in Dianchi Lake basin: dense population, high reclamation rate of land, deficiency of water resource, low forest coverage as well as urban expansion. We need to start the ecological environment development from the above mentioned aspects. During the process, we need set up digital information platform to supervise various data and conduct an overall scientific analysis, and provide decision support for the ecological environment development.

The ecological development of Dianchi Lake basin contains multiple information and data of topography of Dianchi, historical figures, plants and animals, cultural relics and historic sites, museums, rivers, villages, counties and districts. All the above brings a tremendous data, various types including pictures, excel forms, videos and spatial data which forms a complex data. It requires the digital information platform to supervise the data efficiently and reflect some overall information as a strong support for the ecological development of Dianchi Lake basin.

It is necessary to consider composite factors of the history, humanity, geography, plants and animals, landscape as well as rivers in Dianchi Lake basin when establishing the ecological system and environment protection. It is rather difficult to obtain overall information of all these massive data through the traditional RDBMS. However it may bring an abundance of data lack of information. The graphic database organizes and stores the data through graphics, adding or deleting on graphic traversal and search, which can report the mutual relationship between massive data in the form of overall information. We established graphic database and visual query system to analyse the ecological system of Dianchi Lake basin based on targeted modeling and study of query methods on ecological environment characteristics, by adopting graph storage models of Neo4j, a high-performance distributed graphics database.

The establishment of the graphic database and visual query of ecological system of Dianchi Lake basin

It is incapable to obtain overall information through traditional RDBMS due to the massive data resource and volume, various types, complexity of relationship of the ecological development of Dianchi Lake basin.

Neo4j, a high-performance database of NoSQL which does not need fixed table structure, but can easily enlarge and define data format by user in flexible data model designing. It is suitable to utilize Neo4j to handle the massive database for the ecological system of Dianchi Lake basin benefits from its flexible data model, high augmentability, high concurrency reading and writing, efficient storage and accessing of massive data. We therefore use graphic model Neo4j of high performance for storage, data management as well as query for ecological development system data[1].

Data usually includes structured data and unstructured ones. In the same way, various data for ecological system of Dianchi Lake basin also includes the structured data and unstructured data (such as videos, audios and pictures). We utilize graph storage model, taking entity data (excel forms, pictures, videos and audios) as nodes of graphs, the relationship between these entity data as the edges. Making various data of ecological system for Dianchi Lake basin as data samples, establishing relationship between forest species and water quality of rivers, forest coverage rate and water quality of rivers, crops cultivation, fertilization as well chemicals and water quality of rivers, different crops and water quality of rivers to study changes of water quality data of each river, do the cluster analysis of water quality data changes for all rivers drain into Dianchi Lake to establish water environment model and graph data model based on all these relationships, then finally establish graph database through index built on feature extraction.

The emerging of database message retrieval(DBIR) as times requires to cater to many average users who are not know-how of query language and database model. Average users could make message retrieval of database by key words via DBIR[2]. It helps average users get structured information by key words on graphs database query without asking them to have knowledge of complex query language and database model. Through input of key words combining section and expression, it solves the non-ideal query result from uncertain key words or ambiguous context by limited key words query.

Key words query of graph database have high complexity of calculation ,optimizing query measures by building index is widely used in graph database to increase the efficiency [3]. The query process is to decompose the query graphs, build index, merge those qualified into candidate set after filtration, and then testify the candidate set to obtain query results [4].

Dynamic changes of many data for ecological development in Dianchi Lake basin make it necessary a timely updating of database and rebuilding index of many calculations which will consume much time on calculation.

We adopt the incremental subgraph search composite algorithm which could be added into or revised without rebuilding index[5]. To structure the distributed index, to trim and match the nodes via information from graphs before searching, to insert revision without rebuilding index by the time of updating of graphs which could save calculation with an increased search efficiency[6].

In recent years, the application demand for visual query to unfold abstract data increases day by day. Visual query technology became the hot spot and difficulty of study in this field[7].

Because of the complexity and variety of data in the distributed environment of Dianchi Lake Basin, as the data scale is expanded, to avoid the abundance data lack of information, we establish the graph database visual query system which can provide a friendly graphical user interface, visually unfold abstract information, facilitate users study on the data, direct users input to form a visual feedback of the query, and solve the problem that the query results cannot be directly observed, and can reflect the relationship and overall information between various complex data. Technology roadmap of the establishment of the graphic database and visual query of ecological system of Dianchi Lake basin is as figure 1.
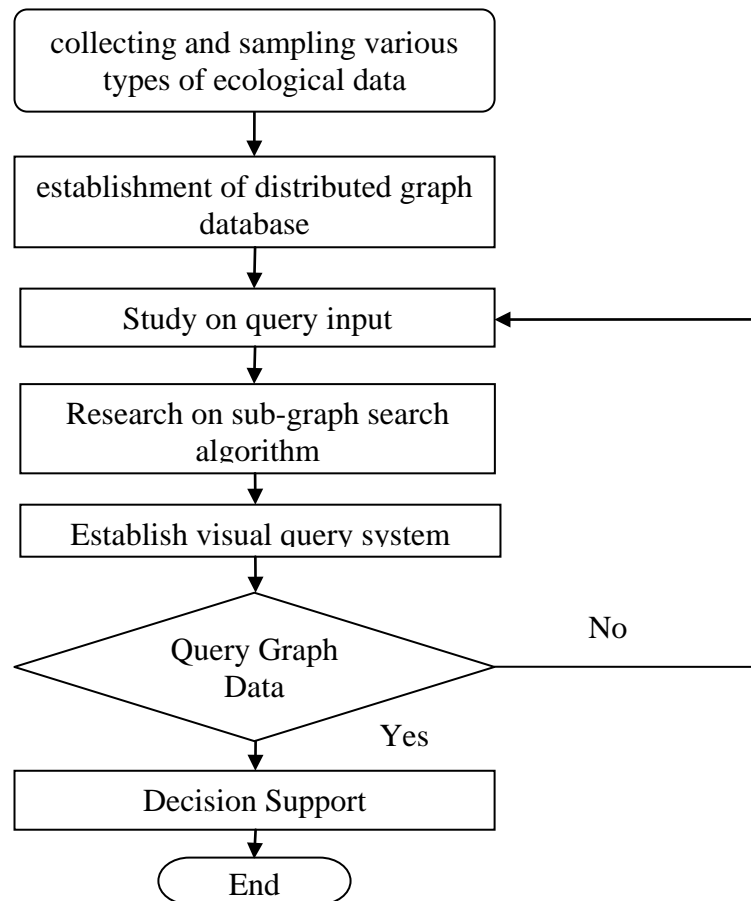
```
┌─────────────────────────────┐
│ collecting and sampling various │
│   types of ecological data      │
└─────────────────────────────┘
               │
               ▼
┌─────────────────────────────┐
│ establishment of distributed graph │
│         database            │
└─────────────────────────────┘
               │
               ▼
┌─────────────────────────────┐
│   Study on query input      │◄──────────┐
└─────────────────────────────┘           │
               │                          │
               ▼                          │
┌─────────────────────────────┐           │
│ Research on sub-graph search │          │
│         algorithm           │          │
└─────────────────────────────┘           │
               │                          │
               ▼                          │
┌─────────────────────────────┐           │
│  Establish visual query system │        │
└─────────────────────────────┘           │
               │                          │
               ▼                No        │
          ◇ Query Graph ◇ ─────────────────┘
          ◇    Data    ◇
               │ Yes
               ▼
┌─────────────────────────────┐
│     Decision Support        │
└─────────────────────────────┘
               │
               ▼
           (  End  )
```

Figure 1Technology roadmap of the establishment of the graphic database and visual query of ecological system of Dianchi Lake basin.

We carried out the graphics database and visual query system simulation experiments of the Dianchi Lake basin ecological development in local area network in small-scale. The results show that the graphics database system is not complex as the updating of traditional relational database system, however with a faster query and high matching accuracy rate. And it is far better than the traditional relational database system in the storage and processing of data for ecological development of Dianchi Lake Basin on its complexity, dynamics and low structure. The comparing results are as follows:

Table1   Comparison between graphics database and relational database

Item \\ Database	Complexity of updating data	Speed of information search	Matching accuracy	Ability to deal with dynamic, complex, the low structure data	Processing power of large-scale data
graphics database	Low	fast	High	Strong	Strong
relational database	High	slow	Low	weak	weak

Conclusion

According to the characteristics of the ecological environment data, we have studied the modeling and query methods, and established the graphics database and visual query system for the analysis of

ecological development in Dianchi Lake Basin. The Neo4j high performance distributed database graph storage model, solves the problem of reconfigurable index when data updates，and solves the query problem of various data sources in distributed environment. It changes the additions and deletions into the traverse to graphics at low complexity and high query speed. By using the incremental sub graph search algorithm, the key words query input method combined with section and expression is adopted to solve the problem of non-ideal query results with a high matching accuracy. By using visual query, a friendly graphical user interface is provided, every step of the incremental query could be generated into visual feedback information. It can solve the problem that the query results can't be directly observed. It can reflect the relationship between massive data and overall information. Simulation results show that this graphics database system in the storage and processing of the complex, dynamic and low structural data of the Dianchi Lake basin is far better than the traditional relational database system. It can effectively analyze and process all kinds of data in Dianchi Lake basin ecological development, and play an important role in the scientific decision-making.

Acknowledgement

References

[1] Yang Shuxin, XuLiping, Xia Xiaoyun, XuHuiqin. Progress of key words study on graphics data [J]，2014,42 (11) :2260~2267.

[2] Yu Suhua, Study on key words retrieval on object level of graphics database [D], Dalian Maritime University, 2013.

[3] Sang Lie, Zhong Ming, Chen Liu, QuJunfeng, Liu Mengchi, FindGrape_A High Efficiency Graphics Database Key Word Search Engine [A]，Section B, the 29th Chinese Academic Papers on Database（NDBC2012），2012：408~412.

[4] Liu Yahui, Liu Chunyang, Zhang Tieying, Cheng Xueqi, Review Study on Graphics Indexing [J].，Shandong University Journal (science edition)，2013, 48（11）：44~52 Tan Wei，Study on Accurate Query and Approximate Query of Graphics Database [D]，Jiangxi University of Science and Engineering，2013

[5] Wu Yong, ZhongZhinong, Jing Ning, Li Xing, Visual Study on Massive Graphics Data，Study on Computers Application [J].，2012, 29（9）：3216~3220

[6] Tan Wei, Yang Shuxin, Study on Approximate Query Calculation of Graphics [J]，Computer Engineering and Designing，2013，34（5）：1700~1705

[7] Zhong Ming, Wang Sheng, Liu Mengchi, An Optimization Method for the Search of Known Terms on Large Scale Graph Data, Computer Research and Development, 2014，51(1) :54~63.