Database Construction of Real Estate Assessment Based on Big Data

Liang Zhou[1,2,a], Liang Shi[3], Sijia Zhang

1Harbin Institute of Technology Shenzhen graduate School, Shenzhen, China

2The Postdoctoral Innovation Base of Shenzhen Assessment and Development of Real Estate Center, Shenzhen, China

3Center for Assessment and Development of Real Estate, Shenzhen, China

aCorresponding author, e-mail: zhoul@lreis.ac.cn

Keywords: Big Data, Real Estate Assessment, Data Warehouse.

Abstract. With the age of big data coming, it provides unprecedented development space to real estate assessment. It's no longer difficult to gain massive, multi-source and multi-scale data. In order to extract effective knowledge and information, the enormous data not only need to be saved reasonably and normatively, but also need to be screening and further mining. This article, which main line is big data, introduces data source and classify of real estate assessment, and then proposes the method and procedure of database creation. It states the effect, value and position of big data for real estate assessment from data integration layer, storage service layer, applied analysis layer and information expression layer. Database of real estate assessment based on big data can solve traditional real estate assessment problems due to lacking of data, and play advantage of data itself through data mining and knowledge discovery at utmost.

The Age of Big Data

According to segment of information processing, big data can be divided into six segments: data collection, data cleaning, data storage and management, data analysis, data manifest and industry application.As is known to all, big data is not just huge data, but analyze the huge data. We can only get intelligent, in-depth and valuable information through analysis. More and more application relate to big data, the nature of those big data, include quantity, speed, diversity and so on, presents the complexity of the constant growing of big data. So, in the field of big data, the analysis method is very important. It's the determining factor for whether the final information valuable or not [7].

Data Foundation of Real Estate Assessment

Data is foundation of real estate assessment, the quantity and quality of data will affect the accuracy and precision of assessment directly.In practice, appraisers always have obsession with experience judgment, such as comparison of price, rate of return, return on capital of building project, cost and composition of construction project. The reason is related to marketization of assessment and related industries.

Take real estate transaction case as example, the main way to get real estate transaction case are (according to reliability): feedback data of real estate agent, consulting data of street shop, listing data of newspaper, query data of famous website and accumulation data of assessment institution. From the point of data source, the data from agent whom help to bring about a deal is the most reliable. And the information is more comprehensive, such as floor, orientations, house type, landscape, ventilation and lighting. It's a great channel for assessment institutions which have agent resources. But there are limitation on coverage rate of agent business, so quantity of information is limited too. And most of assessment institution don't have this kind of resource. So, the most popular way are consult agent of street shop by phone or on site, listing data of newspaper, data of famous website, accumulation data of assessment institution and data sharing by appraisers. The reliability and precision of those data can't meet the qualification. And the data can't reflect individual quality, is just average price, ball park figure and lowest price, so the data isn't

convincing. In assessment dispute, the shortcoming is more obviously, especially for assessment dispute on court. One of difficult problems for appraisers is they don't have reasonable response on reliability and truth of source of the case.If we compared case as basis, only with solid basis, the conclusion will stand firmly. If we only have average price, ball park figure, the conclusion is unstable just as step on several boats at the same time.

At present, the main data sources of real estate assessment are system data of government, system data of enterprise and data from internet. With the development of measure, storage, computer technology, there are more and more abundant collecting methods for real estate assessment data, especially the client data headed by Web 2.0 users increased in a geometrical progression, everybody is a collector and publisher. Traditional attribute data and spatial data become more and more normative, and it's much easier to get the data.

The price data of real estate information is always the key point and difficult point of data collecting. With the age of big data approaching, many real estate related companies cooperate with each other and share the data they have. The development of internet explored the width and depth of data. Through searching, screening and mining the key word on internet, we can acquire information we could not collect in the past. And then summarize the information, we can extract knowledge from information. The real estate assessment based on knowledge is the development direction in the future.

Construction of Data Warehouse

Data base is a subject-oriented, compositive and stable data set which can reflect historical changes. And it's used for supporting management decisions. Data base provides current and historical data which can support users' decision, and these data are difficult to get from traditional operation data base. Data base technique is a general name for all kinds of technique and model which can collect operational data into a centralized environment, and provide effective decision making data.All is for get information users needed more and more quickly and convenient, provide support for decision making.

Data Cleaning and processin. Data preprocessing is the first step for constructing data base, including check consistence of data, handle invalid value and missing value, to ensure accuracy, consistence and integrity of data.

According to realization pattern and scope of data cleaning, it can be divided into 4 kinds [5]:

1. Manual realization. Checking by manual work, it can find all the mistakes if input enough manual labor, material and financial resources, but it's inefficient. It's almost impossible in the case of big data quantity.

2. Application program written special for data cleaning.This method can solve particular problem, but not flexible enough, especially when we need to do cleaning repeatedly. (Generally speaking, it rare could meet the demand for cleaning once) And will lead to complex procedure and heavy work load. And this method can't take full advantage of powerful data process ability of data base.

3. Solve problem in particular application area. For example, seek for abnormal record, clean name, address and post code according to the principle of probability and statistics. It's the field which is researched at present, and the most successful category. Such as business system: Trillinm Software，System Match Maketr and so on.

4. Data cleaning unrelated to particular application area.The research is focus on clean repeated record, such as Data Cleanser Data Blade Module, Integrity system and so on.

Data cleaning for real estate assessment is mainly for real estate data, such as transaction data and price data.Remove error value by statistical analysis technique, such as hypothesis test, significance test and variance analysis. General idea of data reduction is shown in the Fig 1:

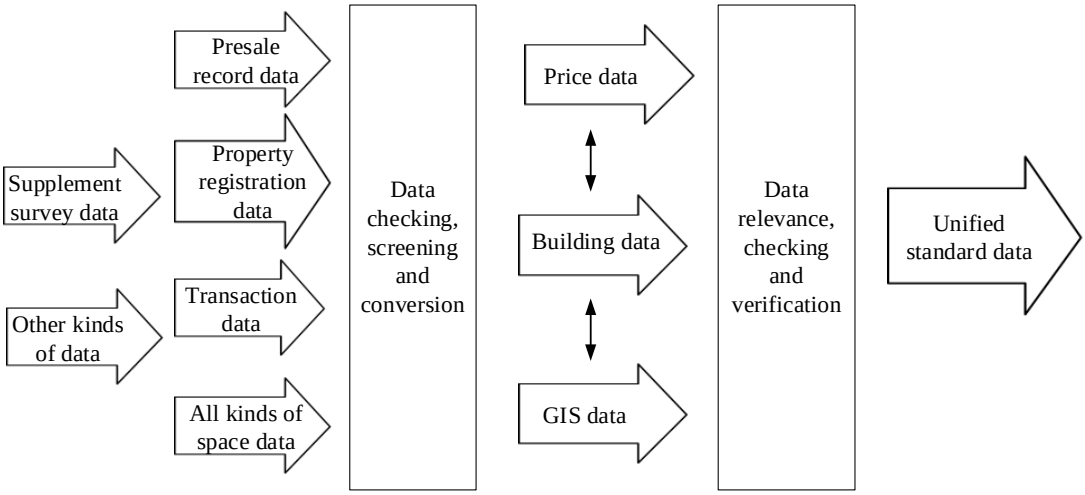

Fig.1 General Idea of Data Reduction

After data cleaning and processing, finally form pyramid model of assessment data as shown in the Fig 2:

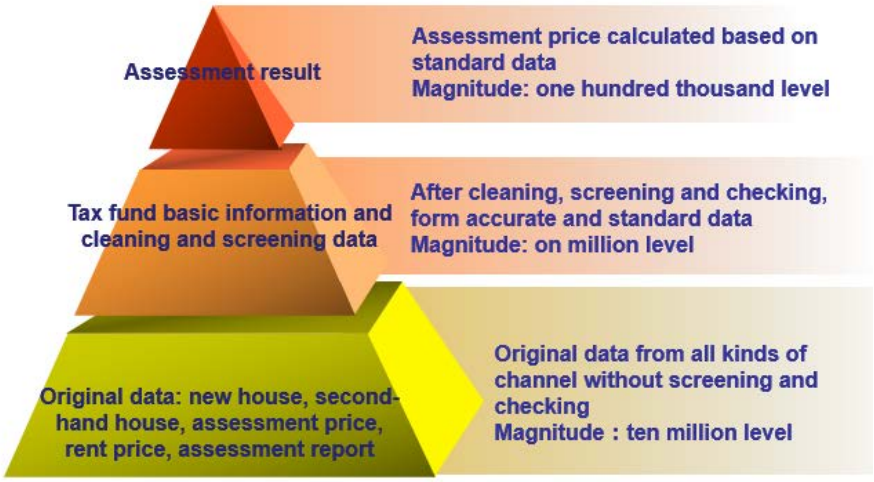

Fig. 2Pyramid Model of Assessment Data

Structure of Data Base. The data of real estate assessment data base is for entire assessment industry, organize and unify all kinds of data sets, and ensure consistency, integrity, efficiency and accuracy of the data. Data base must reflect the change of time, pay attention to historical data. Most of the charts include time attributes, and can receive new data regularly. The structure of real estate assessment data base is shown as figure 3. Data source is the basis of data base, is the source of entire system. It includes interior information and external information. Interior information is interior business data and related documents. External information is market survey and analysis information and all kinds of documents.These data source can be constituted by different type of structure data documents, such as large scale relational data base - Oracle, middle and small scale relational data base - SQL Server, desktop data base - Access and database file - Word, Excel.

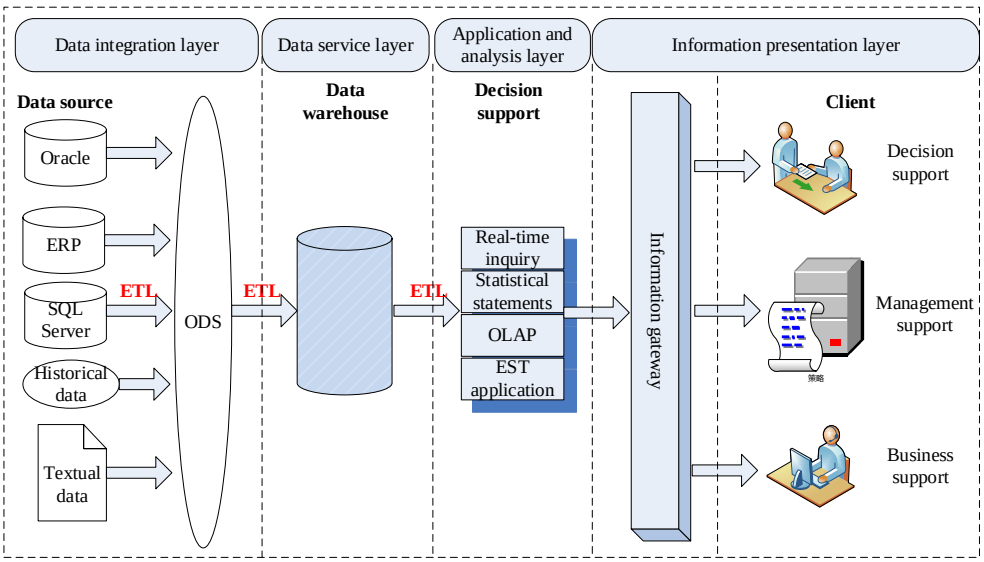

Fig. 3 The Structure of Real Estate Assessment Data Warehouse

Storage and management of data is the core of entire data base system, it is in charge of interior maintain and management of the data base.The interior maintain of data base include construction of data structure, data manipulation, data maintain and control and data service. The management of data base include data security, file, back up, maintain and recovery.Literature [3] introduced construction of real estate assessment data base detailed.

Client application is for end users, it includes front tool and application. The front tool mainly includes all kinds of analysis tool (OLAP), reporting tool, query tool, data mining tool and application for data base or development of data mart.Based on internet data, we can construct knowledge data base.Different from traditional relational data base and spatial data base, knowledge data base will screen, clean and well-ordered save data, extract information and knowledge which people don't know but useful through data mining method. According to the different data type, the knowledge we can mine is different too, including geography, price, transaction and users' decision. So, we can construct different subject data base.

Application of data base. After we finish construction of data base, data integration layer provides unified entrance for the entire system, receives massive data sent from data collection terminal quickly, cut down the time for data collection, and reduce concussion to application system from data collection. At the same time, it can collect data cross system and from several data source, improve the reliability and consistency of data collection.The data in data base organize by subject according to specific model, every analysis subject matches fact table and dimension table in the data base.Construct multi-dimensional analysis models by data in the data base according to the demand of real estate assessment, save by ROLAP or MOLAP, so that users can share, access and analyze data from any angle and any detail.Finally, provide information and analysis result in data base to end users according to users' analysis demand.The integrative organizational structure of real estate assessment data base is shown as Fig 4:

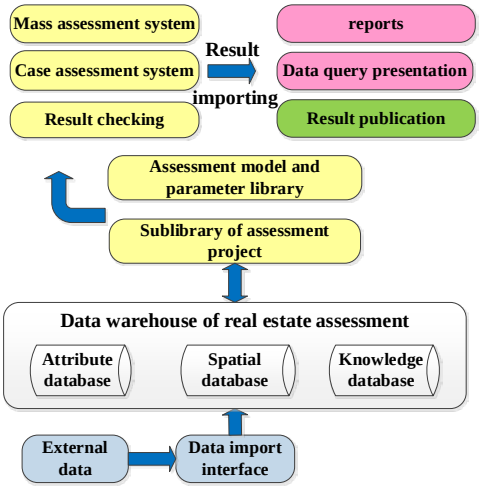

Fig. 4 The Integrative Organizational Structure of Real Estate Assessment Data Base

Conclusion

Real estate assessment is a process of data mining, quantity and quality of the data will impact result of real estate assessment directly. In the past, the problem of lack of samples and false data restricted the accuracy and precision of the results. With the age of big data coming, it solves the problem of data lacking, the new problem is there are too many data. The primary issue is organization, storage and cleaning of massive data. With the development of distributed, cloud technology and data base, the new problem for big data is how to make good use of the data further.The data base can solve problems of storage, classification, mining and utilization, lay the foundation for the development of real estate assessment. In the future, construction of data model according to subject and data mining based on industry characteristic will be the research focus for data base.

References

[1] Snijders, C. &U. Matzat& U.D. Reips. “Big Data”: Big gaps of knowledge in the field of Internet science. International Journal of Internet Science. 2012(7), 1-5.

[2] Viktor Mayer-Schönberger . The era of big data - big changes live, work and thinking . Zhejiang People's Publishing House. 2012(12).

[3] J.J. Geng, H. Zhang. Construction of GIS-based real estate mass appraisal database - A Case of Shenzhen. Sensing Technology and Application. 2012(3).

[4] Song Ji. Mining technology research and application in real estate information analysis system based on data warehouse data. East China Normal University. 2007.

[5] Xinag Xu, Jie Mao . Application of data cleaning techniques in software development. Computer era. 2004(8)，25-26.

[6] F.X. Yang, Y.C. Liu, Z.H. Duan. Summary of data cleaning. Application Research of Computers. 2002，19(3)，3-5.

[7] J.H. Zhao. Opportunities and Strategies Challenges faced by libraries Big Data era [EB/OL]. http://www.doc88.com/p-9768750532253.html, 2012.