Design and Research of Statistical Analysis System Based on Business Decision Field

Xiaolian Li *, Dexin Qiao, Yu Ding, Ying Shi, Wei Guo, Daiming Wei
Research Institute of Petroleum Exploration & Development, Beijing 100083, China.

* Corresponding author. Tel.: +8613131455889; email: dingyu89@petrochina.com.cn
Manuscript submitted May 13, 2020; accepted July 8, 2020.
doi: 10.17706/jsw.15.6.172-180

Abstract: With the gradual deepening and expansion of the scale of enterprise information construction, it is more and more required that the system can break the business barriers and data barriers, and realize a new software design idea of data sharing, technology sharing and collaboration. At present, the data statistics system of large-scale energy enterprises has been unable to meet and adapt to the increasingly complex and huge data analysis demand group. There are many kinds of calculation caliber and groups, and the data processing capacity has multiplied. The original system has been overwhelmed. Therefore, a set of functions is needed to support the comprehensive statistical business needs of the group headquarters, integrate discrete statistics, and provide integrated big data statistical business management; finally, the functions of data collection, data verification and data presentation of the comprehensive statistical business of the company are realized, so as to improve the level of statistical business data query and data analysis.

Key words: Intelligent analysis, big data collection, system design, business decision-making.

1. Introduction

With the increasingly fierce market competition, management personnel at all levels of the company must be able to timely respond to market changes and make correct business decisions. Whether it is the decision of investment, the decision of resource allocation, or the decision of controlling production costs and costs, the market decision of market development and product promotion, all are closely related to enterprise statistics. The analysis and evaluation of statistics and information is an effective tool to supervise all aspects of the company's operation and management, and is also the basis and basis for strategic decision-making and management. Therefore, we should further strengthen statistical management and improve the statistical information system. At the same time, we should organize statistical monitoring, early warning, analysis and evaluation carefully, so that the company can avoid mistake in decision-making, supervise and control the implementation of decision-making, and ultimately guarantee the company to obtain greater economic benefits. In 2007, Yu Dan, a senior statistical analyst, studied the comprehensive reliability evaluation methods and data filling algorithms in the reliability analysis of complex systems [1]. In the face of imperfect data validation and simple analysis dimensions, a comprehensive statistical system suitable for general industry [2]. There is also a mixed structure of C/S and B/S to build an analysis and statistics system model [3]. The above research on statistical analysis often has data deviation and analysis angle deficiency in the practical application of enterprises, which leads to

low practical rate. In this paper, based on the actual project experience, it will be studied in depth, and combined with big data technology. Using the experience accumulated by the previous generation system, the bottom layer of business architecture is designed innovatively. At the same time, the new ideas of DMS model and data analysis framework design are introduced to improve the accuracy of data calculation and data verification services, and give full play to the role of business architecture in standardizing business. Finally, improve the efficiency of index and dimension matching, prevent and control risks, and realize data collaborative sharing.

2. Core Standards for System Construction

2.1. Statistical Index System

For the statistical system, the construction of a set of index system suitable for the enterprise itself undoubtedly lays a foundation for the successful construction of the system [4]. Then according to the theoretical basis such as enterprise's own maturity and economic evaluation [5], the study will start from two aspects of the system construction.

Dimension type：Firstly, from a single perspective, the complete set of dimension members is obtained. Usually there is a hierarchical summary calculation relationship where dimension members can dynamically increase or decrease without affecting the calculation relationship. Moreover, completeness does not necessarily mean that they can be added together. It depends on the basic indicators of the specific combination, such as the catalogue and the output combination, which cannot be added together, and when combined with the sales amount. Secondly, it is viewed from multiple perspectives and the addition of dimension member values does not have any business implications. Further labeling is required if there is a calculation or verification relationship between dimension members of other dimensions. An integrated dimension can also be essentially a set of indicators. Finally, when a dimension has more than one child dimension and the child dimension has different dimension types from the parent dimension, the type of its child dimension needs to be further labeled.

Indicator type：Indicators are concepts that describe the quantitative characteristics of a population. The basic composition of an index consists of two elements: the name of the index and the value of the index. It reflects the characteristics of two directions: the prescriptive nature of things and the prescriptive quantity of things. In some cases, the indicator may have multiple units of measurement (e.g., tons and barrels of crude oil) or no units of measurement (e.g., number of equipment, in which case the specific unit of measurement is determined by the combined dimension members). Indicators are divided into basic indicators and derived indicators. On this basis, this study also defines the derived class index, which is derived from one basic index and multiple dimensions (members). Derived indicators can be used directly by the system, or more granular indicators can be derived based on derived indicators and dimensions.

2.2. Usability and Extensibility Construction

The development of information support is changing rapidly. At the same time, with the advancement of enterprise information construction, systems in various business areas are gradually established and matured. Therefore, we need to plan the usability and extensibility of the system in depth so that it can be flexible and change with the increasing volume of business. It can also interactively dock with other business system interfaces through a unified standard.

The system uses a microkernel Scalable Mode [7]. In the design of comprehensive statistical system, the most challenging is the flexibility of business entity models. Indicator system, system report, statistical report and calculation formula all require flexible configuration. A system with a large amount of configuration, whether it is development or maintenance, is very heavy work. The most basic design

principle is to avoid overlapping business complexity with technical complexity. It is a good choice to use a streamlined system core and expand and improve step by step. As shown in Fig. 1.

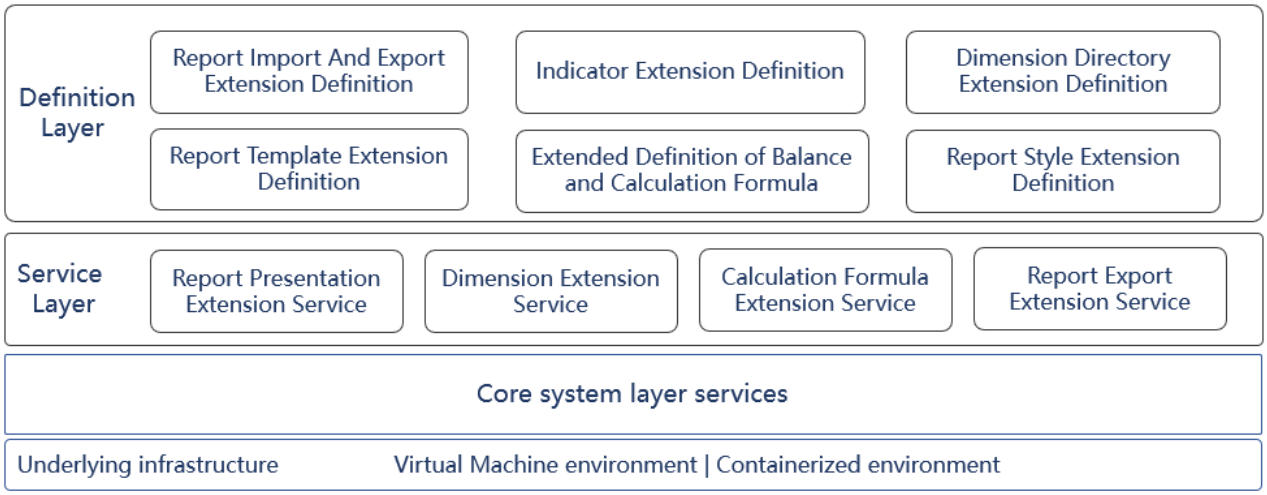

Fig. 1. System core and extension model diagram.

The core system is the smallest set to complete the core business functions of the system, even providing only the basic definition, processing and operation functions. Although simplified, the functional requirements are complete. The functionality of the core system can be applied to most common business needs.

The first level of core system expansion can be in four directions, dimension, formula, presentation, import and export. This level of expansion is necessary for the reporting business of the headquarters system, and also needs to be achieved before the system comes online. Dimensional business conceptual model is relatively simple, but many special and regular business phenomena in system implementation can be reflected by changes in dimensions. At this point, the dimension is no longer just a complete division of indicators. For example: the current period, the same period, the previous period; subitems, parent items, values; this month, the cumulative end of this month, and so on. From the perspective of system expansion, it is natural to expand from the dimension concept in form. Formulas, if more than four operations, can be generalized at this level to begin to expand. Special presentation needs are also placed on the periphery to expand and separate from the kernel. Similarly, Excel import and export are implemented on the periphery.

3. System Architecture Model Design

The system architecture model is the overall description of the software system organization, structure part and system containing interfaces. From the top-level user to the operating-side component design, then to various service interfaces, and finally to interact through database access. The specific architecture diagram is shown in Fig. 2.

Web User Interface Components: Interface components display data to users, get and validate data they enter, and interpret user actions they perform on the data. Interface components should provide visual hints and validation information when accepting user input. Capture user events and change the way data is displayed. Limit the types that users can enter, such as a field value that can only enter numbers. The validity of performing data entry, such as restricting the range of values that can only be entered in a specific field. In addition, the interface component should fetch data from the application's business components and perform data formatting when rendering data.

User interface process components: User process components are bridges between user components

and business components, which are passed through business entities as parameter objects. Process components are components that maintain the state of business processes and user interface components and ensure user-friendly interaction with the system.

Business Workflow： Business workflows enable process flexibility through configuration rather than hard coding.

Business Logic Components: Business logic components implement the business logic of the system, mainly to implement business rules and perform business tasks.

Service Proxy: When a system requires the functional implementation of an external service, a service proxy can encapsulate a method of invoking the external service. You can also convert the format in which the data is delivered. Even if the format of the external service data changes and only the service proxy portion needs to be modified, the format of the data is fixed for the system [8].

Service Interface: In order to provide services to the outside world, business components need to establish service interfaces to support various kinds of communication, etc. In some cases, service interfaces can also represent business interfaces.

Data Access Logical Components: Access to a data store is required during business processing, and the logical and physical structures of entities need to be strictly separated for better configuration and maintenance.

Business entity components: A business entity represents a real-world business entity and is typically a data structure, such as an investment project, where entities are generally passed between components as parameters.

Operation Management and Security Management: Operation Management and Security are mainly to manage exception handling, verify that users perform specified tasks, and so on.

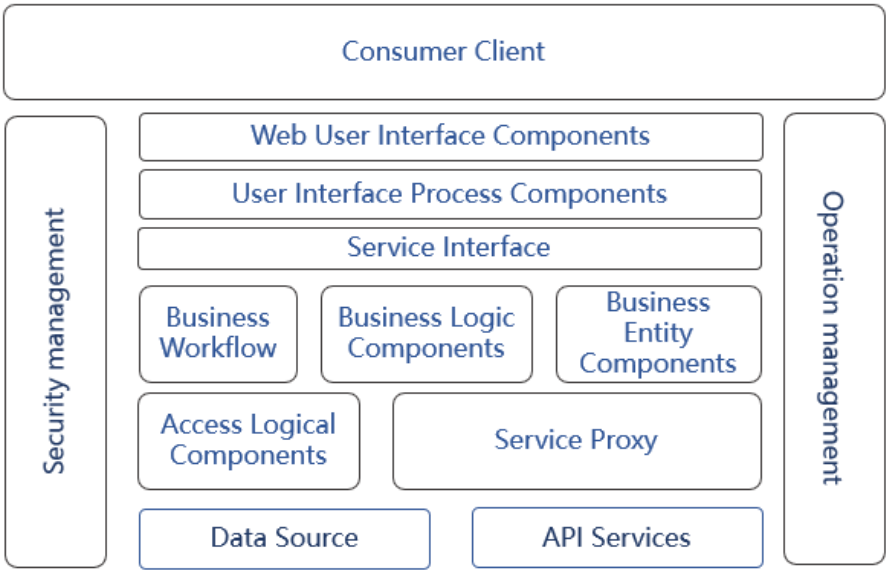

Fig. 2. System architecture model.

4. The Key Technologies

4.1. DMS Model Design

The most important form of presentation in statistical analysis is to calculate in an intuitive report mode. Therefore, in designing a dynamic report, we need to focus on all elements in the data area.

The data area is a data matrix of M rows and N columns. The values of M and N are determined by the

definition of the subject object column. No matter how the main column and object column are combined by dimensions, they will form corresponding tree directories. M is the number of leaf nodes in the main column directory tree, and N is the number of leaf nodes in the object column file tree. Formally, the content of the data area is passive, and there is no element to be extracted, but a key problem is brought out: how to identify the data of the cells in the data area, and an important dimension member set (DMS) is brought out.

DMS is constructed to compress business data storage to the maximum extent and express normal business meaning with the simplest data. For example, "sales volume of special gasoline" and "sales volume of special gasoline" have the same meaning in business, so the system solves this problem by sorting and recombining the leaf nodes.

DMS can, but is not limited to, the data used to identify cells. It combines with an indicator to form IDMS (index qualified by dimension member set), which is the unified identification of granularity statistical data under the indicator system and the basis of data access, summary and formula calculation [10].

4.2. Data Analysis Model Design

This section provides formula definitions for system templates. Including calculation formula and balance formula. Provides a formula definition for the statistics template. Including summary formula and dimension formula.

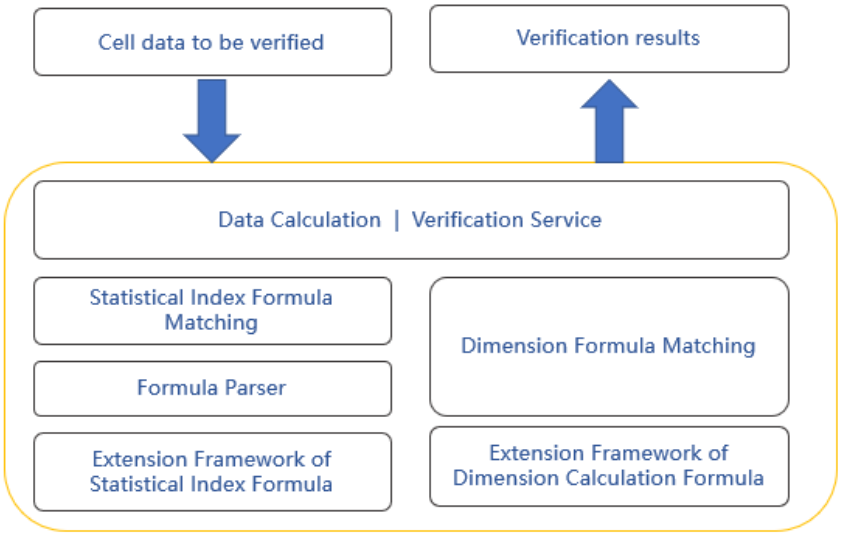

Fig. 3. Data analysis framework model.

In particular, it needs to be noted that indicator formula matching and dimension formula matching. First of all, if all the members in the indicator formula match successfully, use the formula parser to query the relevant indicator formula according to the relationship set between the indicator and dimension members in the distribution report. Indicator formula types include calculation formula, summary formula and balance verification formula. Secondly, if the dynamic members in the dimension formula match successfully, use the formula parser to query the relevant dimension formula according to the customized dynamic dimension in the distribution report. Dimension formula types include: cross cycle ditto, weighted dimension (corresponding indicator), child parent item dimension and extended attribute dimension.

Get the current report formula collection from the report formula database according to the current report number. Obtain the corresponding data in the current report according to the report members (indicator + dimension member set) in the formula. According to the formula in the interface, the Eval function in JavaScript is used to calculate the formula value in the interface.

4.3. Dimension Extension Design

The definition of system dimension is divided into two categories: general dimension and extended dimension. The dimensions defined in the indicator system are general dimensions. The definition of general dimensions and their members can be completely completed at the business level. The system processes the general dimensions in a unified way. Extension dimension is a kind of special dimension, whose members and system processing methods are different, so it needs to be expanded at the technical level. Examples of extension dimensions include: current period, same period, previous period; current month, cumulative up to current month; value, proportion; sub item, parent item, value, etc. These dimensions do not need to be defined in the indicator system and are provided by the system. The addition and change of these dimensions need to customize and develop corresponding functions.

The definition of dimension mainly affects three links of system report: report definition, data filling and data loading. Accordingly, the system defines three abstract interfaces to isolate the corresponding concrete implementation. As shown in Fig. 4.

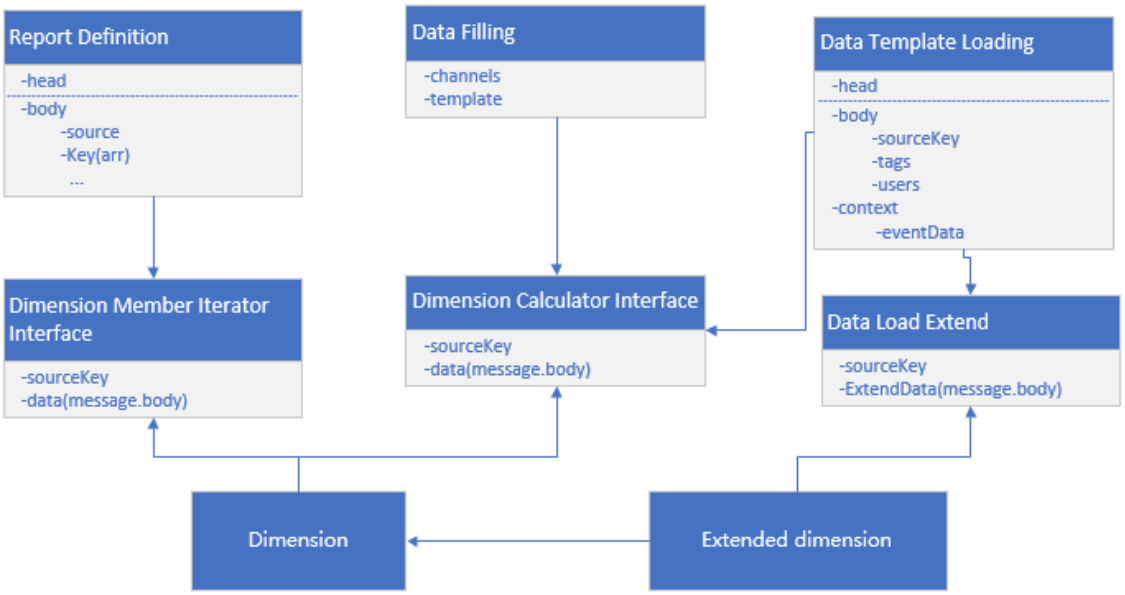

Fig. 4. Dimension expansion diagram.

Taking data loading as an example, the system uses the factory (interface and constructor) pattern to isolate the impact of changes in extended dimensions on the system. The change of requirement for extension dimension only affects the configuration metadata of report definition. The requirement for the development of special dimension of system report can be met by adjusting metadata and deploying new processing components of extension dimension.

5. System Implementation

After the completion of the project, the system has been deployed on the server (windows 2012r version) in the form of web application and included in the real-time monitoring of operation and maintenance. After testing all data sources, it is consistent with external integration interface and page interaction and requirements. The main function module of the system is shown in Fig. 5, the dimension member set model is shown in Fig. 6, and the data balance and verification are shown in Fig. 7.

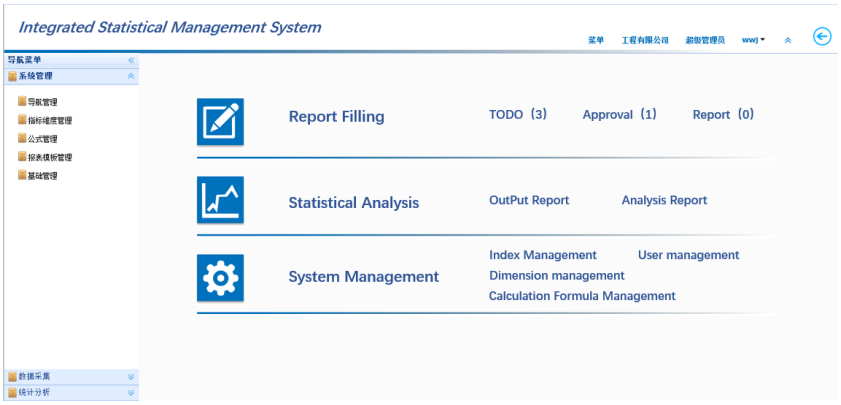

Fig. 5. Main function area of the system.

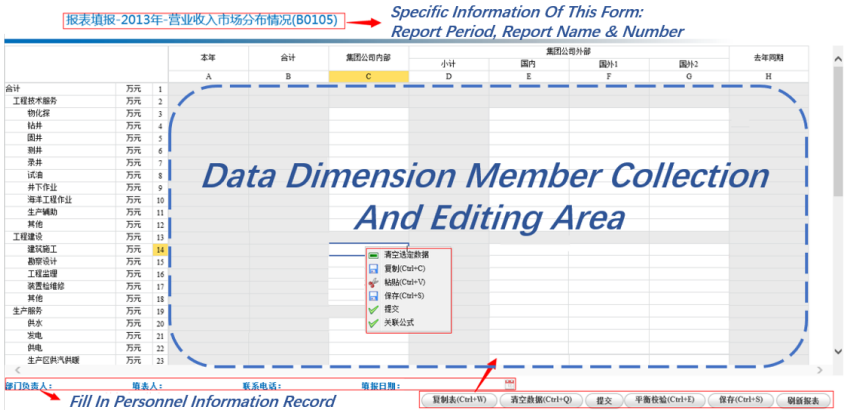

Fig. 6. Dimension member collection model.

Fig. 7. Detail page after data source balance verification

6. Conclusion

In view of the current business situation that enterprises urgently need to make decision-making statistical analysis on business decision-making. Based on the model of detailed statistical index system, this paper introduces the core standard of system construction, which is a guidance method for system implementation. Through the key core techniques such as DMS model design, formula model design and dimension extension design, the business accuracy and advanced nature of the integrated statistical

management system are promoted, and the system's robustness and high data utilization requirements are met. At present, the system has been put into operation. It has been proved that the system designed and built with the core construction standards and system architecture described in this paper can not only highly adapt to all kinds of business needs encountered in business decision-making. Moreover, it can realize the rapid output of data, and improve the success rate and construction speed of the project. At the same time, it has good expansibility and applicability, which provides important practical guidance for the construction of other systems in the future.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Xiaolian Li is responsible for the implementation and planning of the project. In this study, Dexin Qiao developed the overall business plan, and gave detailed problem solutions face to face. As the main program design director, Yu Ding provides overall program guidance and technical support for the project, and has made great efforts in the paper writing. Ying Shi provides effective technical support as the development manager of the project. Wei Guo is responsible for the combing and post-operation and maintenance of business plans. Daiming Wei provided effective support for data correction in this project. Thank all team members for their contributions to this project.

References

[1] Yu, D., Li, X. J., Jiang, N. N., & Yang, J. (2007). Some statistical problems and progress in reliability analysis of complex systems. *Journal of Systems Science and Mathematical Sciences*, 68-81.

[2] Zhu, P. (2017). Research on integrated statistical management information system. *Computer Knowledge and Technology*, *13(10)*, 72-73.

[3] Gong, B. J., Bai, D. Y., & Han, J. (2008). Research on the comprehensive statistical system of coal enterprises based on the mixed structure of B / s and C / s. *Statistics and Decision*, *22*, 187-188.

[4] He, T. (2019). Thinking about the improvement of statistical work in the era of big data. *China's Collective Economy*, *20*, 158-159.

[5] Arpan, M. (2016). Research on the innovation engine model of enterprise management innovation capability evaluation, *5(3)*.

[6] Wang, Y. F. (2019). The application of statistics in modern enterprise management. *Proceedings of the 2019 5th International Workshop on Education, Development and Social Sciences* (pp. 173-177).

[7] Qian Zhenjiang, Lu Liang, Huang Hao. Research on the formal design of multithreading mechanism in microkernel architecture [J]. Computer science, 2013,40 (04): 136-141 + 163

[8] Sun, F. (2014). Research and implementation of integrated statistical management information system. Changjiang University.

[9] Zhang, X. H., & Wang, P. K. (2000). User authority management of large information system. *Computer Application*, *11*, 35-36 + 39.

[10] Pan, F. C, & Zhang, S. J. (2005). Research on XML based intelligent report generation tool. *Minicomputer System*, 134-138.

Xiaolian Li has completed his bachelor's degree in computer science and technology in China University of Petroleum from 2006 to 2010, and he received his master's degree in computer science and technology in China University of Petroleum from 2010 to 2013. Since then, he worked as a software development engineer in the Institute of computer application technology of Research Institute of Petroleum Exploration & Development in 2013. At present, he is engaged in cloud Computing and Information Intelligence Technology Research.

Dexin Qiao is a senior lecturer and IT program leader at Computer Application Technology Department, PetroChina Research Institute of Petroleum Exploration & Development, Beijing, China. He has published number of research papers on ensemble learning. His expertise and research interests include software design, ensemble learning and knowledge discovery.

Yu Ding received his master's degree in computer science and technology from Griffith College, Ireland, in 2015. Since then, she worked as a software development engineer at the Institute of Computer Application Technology in 2016. At present, she is working in the software architecture design of Research Institute of Petroleum Exploration & Development.

Ying Shi received the master's degree in computer science and technology from the Research Institute of Petroleum Exploration & Development in 2017. Since then, he has been engaged in software application research for a long time, with rich experience in development and project management. At present, he holds the position of development manager in the computer application research institute.

Wei Guo began to work in 2007 and graduated from the School of Computer Science of China University of Petroleum in 2010. He mainly engaged in system design, participated in and built large-scale comprehensive statistical system. At present, he works as a team leader in the computer application research institute. He is interested in big data mining and statistical analysis.

Daiming Wei has completed his bachelors in computer science from 2009-2013 in China. He worked as a software development engineer at Computer Application Technology Department from 2018. His research interests include software design, ensemble learning and software development.